

On spatial coalescents with multiple mergers in two dimensions

Benjamin Heuer and Anja Sturm¹

*Institute for Mathematical Stochastics
Georg-August-Universität Göttingen
Goldschmidtstr. 7
37077 Göttingen, Germany*

Abstract

We consider the genealogy of a sample of individuals taken from a spatially structured population when the variance of the offspring distribution is relatively large. The space is structured into discrete sites of a graph G . If the population size at each site is large, spatial coalescents with multiple mergers, so called spatial Λ -coalescents, for which ancestral lines migrate in space and coalesce according to some Λ -coalescent mechanism, are shown to be appropriate approximations to the genealogy of a sample of individuals.

We then consider as the graph G the two dimensional torus with side length $2L + 1$ and show that as L tends to infinity, and time is rescaled appropriately, the partition structure of spatial Λ -coalescents of individuals sampled far enough apart converges to the partition structure of a non-spatial Kingman coalescent. From a biological point of view this means that in certain circumstances both the spatial structure as well as larger variances of the underlying offspring distribution are harder to detect from the sample. However, supplemental simulations show that for moderately large L the different structure is still evident.

Keywords: spatial Cannings model, coalescent, Λ -coalescent, spatial coalescent, two dimensional torus, limit theorems

2010 MSC: 60J25, 60K35, 92D25, 92D10

1. Introduction

The *goal* of this article is to study the genealogies of a sample of individuals from a spatially structured population when the variance in the number of each individual's offspring is relatively large. Larger variances in the offspring distribution are thought to arise, for example, due to particular reproduction mechanisms of various species leading to the existence of few individuals with many offspring (Eldon and Wakeley, 2006), and also due to recurring selective sweeps (Durrett and Schweinsberg, 2004, 2005).

The space is structured into discrete sites of a graph G with a colony of a fixed number of individuals at each site in G as well as migration between sites. As the underlying population models we introduce spatial Cannings models, which are extensions of the stepping stone model with general Cannings type offspring distributions.

¹Corresponding author: Email asturm@math.uni-goettingen.de

The genealogies are modeled by coalescent processes that code for the ancestral lines of the individuals in a sample from the present day population backwards in time. Coalescence -referring to a merger of ancestral lines- occurs when a common ancestor of various individuals is reached. We consider coalescent models that are appropriate if the population at each site of G is large. Our special focus in the analysis lies then on models where the number of sites $|G|$ in the graph is finite but also large, more precisely, we choose G to be a large two dimensional torus.

The special case when one considers only one site (the non-spatial situation with $|G| = 1$) leads to the classical *Kingman coalescent* with only binary mergers provided that the variance of the offspring distribution stays bounded in some sense as the population size N tends to infinity. This coalescent process has been well studied since its introduction by Kingman (1982a; 1982b), see for example Wakeley (2009) for an overview.

Genealogies for populations with a larger variance in the number of offspring have been studied by Mathematicians and Biologists in the more recent past. In this case, the genealogy is described by the coalescent with multiple mergers, the so called Λ -coalescent, which was independently introduced by Pitman (1999) and Sagitov (1999). (More generally, one can even consider coalescents with simultaneous multiple mergers, the so called Ξ -coalescents, but we will focus here on the subclass of Λ -coalescents.) Here, Λ is a finite measure on $[0, 1]$. When there are currently b distinct ancestral lines then any collection of k ancestral lines coalesces and thus merges into one new ancestral line at rate

$$\lambda_{b,k} := \int_{[0,1]} z^{k-2} (1-z)^{b-k} \Lambda(dz) \quad (1)$$

with $2 \leq k \leq b$, $k, b \in \mathbb{N}$. We formally extend this definition by setting $\lambda_{b,k} = 0$ for $b = 1$ or $b = 0$, $k \in \mathbb{N}$. The Kingman coalescent, that is appropriate for populations in which the offspring variance is not so large, corresponds to the case $\Lambda = \delta_0$, the delta measure at 0. In this case $\lambda_{b,k}$ is only nontrivial if $k = 2$ and we obtain $\lambda_{b,2} = 1$ so that the total coalescence rate is $\binom{b}{2}$. This means that only binary mergers of pairs of ancestral lines are possible and happen at rate 1 per pair of ancestral lines.

In order to extend this to a *spatial or structured coalescent* on a graph G we imagine that ancestral lines migrate independently from site x to site y in G with rates $p(x, y)$ but coalesce independently at each site according to the rates given in (1). The formal definition of this *spatial Λ -coalescent* can be found in Section 2 and in particular in Definition 2.1. In the case of Kingman coalescence at each site the resulting process is often simply called the structured coalescent, which was derived rigorously from a forward population model by Herbots (1997) but already studied earlier by Notohara (1990) and since then by many others, see for example Donnelly and Kurtz (1999) or Greven et al. (2005). The case when coalescence at each site takes place according to a Λ -coalescent, was included in a setting considered by Donnelly and Kurtz (1999), and was studied in more detail by Limic and Sturm (2006).

After introducing the spatial Cannings model in Section 3 we show that spatial Λ -coalescents arise when one considers the genealogy of individuals sampled in those spatial Cannings models, provided that the number of individuals at each site is large, see Proposition 3.1 for a precise formulation of the result. This result is a rather straightforward generalisation of the work of Herbots (1997) on the derivation of the structured coalescent combined with the work of Möhle and Sagitov (2001) on the derivation of Λ -coalescents from non-spatial Cannings models, which is included here for completeness.

In Section 4 we consider the spatial Λ -coalescent on the two dimensional torus $T^L = [-L, L]^2 \cap \mathbb{Z}^2$ with side length $2L + 1$. In our main result, Theorem 4.1, we show that as L tends to

infinity, and time is rescaled appropriately, the partition structure of spatial Λ -coalescents of individuals sampled far enough apart converges to the partition structure of a non-spatial Kingman coalescent. This extends in particular the work of Limic and Sturm (2006), which dealt with the case $d \geq 3$. The two dimensional case considered here is biologically the most relevant. It also needs to be treated differently mathematically.

Since the information in a genetic sample only depends on the partition structure of the genealogy of the sampled individuals as well as on some (independent) mutation process along the ancestral lineages, the distribution of quantities that can be read off a sample of genes, such as allele frequencies, will be the same if the partition structures of the genealogies are the same. (Note however that the time rescaling that is used in Theorem 4.1 depends explicitly on L .)

From a biological point of view, Theorem 4.1 thus implies that spatial structure as well as larger variances of the underlying offspring distribution are harder to detect from a sparse sample on a large space. However, we also conduct simulations that show that for moderately large L the different structure is still evident in the frequency spectrum of a sample. We postpone the detailed description of relations of our main limit result to the existing literature as well as further discussion of the analytical as well as simulation results to Section 5.

2. The spatial Λ -coalescent

In this section, we rigorously define the spatial Λ -coalescent. We start with introducing some necessary notation. First, in order to not only describe the number of ancestors of a sample of size n at a given time, but the structure of the genealogy, it is convenient to consider the coalescent as partition valued. We label the sampled individuals from 1 to n and want to describe the corresponding n -coalescent, first in a non-spatial setting.

Here, we consider as the state space \mathcal{P}_n , the partitions of $[n] := \{1, \dots, n\}$. We call the elements of a partition $\pi \in \mathcal{P}_n$ blocks and may represent π uniquely by

$$\pi := (B_1, B_2, \dots), \quad (2)$$

where $B_l \subset [n]$ with $\sum B_l = [n]$ are the blocks indexed by the order of their minimal element (with the convention that $\min \emptyset = \infty$). Each block represents the common ancestor of the individuals contained in this block. Initially, the process is generally started with all singletons, so in the state $(\{1\}, \dots, \{n\}, \emptyset, \dots)$. At each coalescence event an appropriate collection of blocks is merged into one new larger block.

Λ -coalescents can more generally be defined on the state space \mathcal{P} , the partitions of all of \mathbb{N} with representations as in (2). Due to the exchangeability of all blocks an n -coalescent is obtained from this coalescent started with infinitely many individuals if one restricts the attention to the partition structure induced on a subset of \mathbb{N} of size n .

For describing the spatial distribution of ancestral lineages of samples of size $n \in \mathbb{N}$ we consider as the state space for the spatial Λ -coalescent labeled partitions \mathcal{P}_n^ℓ of $[n]$. This is the set of partitions of $[n]$, for which each block carries a label in G specifying its location. To make the notation more precise, we will represent an element $\pi^\ell \in \mathcal{P}_n^\ell$ uniquely by

$$\pi^\ell := ((B_1, \zeta_1), (B_2, \zeta_2), \dots), \quad (3)$$

where the sequence of blocks is ordered as before and $\zeta_l \in G, l \in \mathbb{N}$ are the labels of the ordered blocks. Set $\zeta_l = \partial \notin G$ if $B_l = \emptyset$.

If we are interested in the genealogy of samples of infinite size we consider labeled partitions \mathcal{P}^ℓ of \mathbb{N} . Elements of \mathcal{P}^ℓ are also uniquely described by (3) with $B_l \subset \mathbb{N}$, where $\sum B_l = \mathbb{N}$ and $\zeta_l \in G$. For π^ℓ in \mathcal{P}_n^ℓ or \mathcal{P}^ℓ of the form (3) we will write π for the corresponding unlabeled partition $\pi := (B_1, B_2, \dots)$ in \mathcal{P}_n or \mathcal{P} respectively. We will denote by $\#\pi = \#\pi^\ell$ the number of (nonempty) blocks in the partition, which may be finite or infinite. Also write $\#\pi_x^\ell$ for the number of blocks with label $x \in G$.

For any element $\pi^\ell \in \mathcal{P}_n^\ell$ or \mathcal{P}^ℓ with $n \geq m$ define $\pi^\ell|_m \in \mathcal{P}_m^\ell$ as the labeled partition induced by π^ℓ on \mathcal{P}_m^ℓ . We equip \mathcal{P}_n^ℓ with the metric

$$d_n(\pi^{\ell,1}, \pi^{\ell,2}) = \sup_{m \in [n]} 2^{-m} 1_{\{\pi^{\ell,1}|_m \neq \pi^{\ell,2}|_m\}}, \quad (4)$$

and likewise \mathcal{P}^ℓ with the analogous metric d , where we replace $[n]$ with \mathbb{N} . It follows that $(\mathcal{P}_n^\ell, d_n)$ and (\mathcal{P}^ℓ, d) are both compact Polish spaces (for finite G), and that $d(\pi^{\ell,1}, \pi^{\ell,2}) = \sup_n d_n(\pi^{\ell,1}|_n, \pi^{\ell,2}|_n)$. Analogously, we defined a metric on the unlabeled partition spaces \mathcal{P}_n and \mathcal{P} by omitting all the superscripts ℓ in the above.

We are now ready to rigorously define the spatial Λ -coalescent. Let $D(\mathbb{R}_+, E)$ be the space of right continuous functions from \mathbb{R}_+ into a metric space E that also have left limits. We equip the space $D(\mathbb{R}_+, E)$ with the Skorohod topology, see for example Chapter 3 of Ethier and Kurtz (1986) for details.

Definition 2.1. *The spatial Λ -coalescent Π^ℓ with parameter $\Lambda = (\Lambda_x)_{x \in G}$ where Λ_x are finite measures on $[0, 1]$ is a $D(\mathbb{R}_+, \mathcal{P}^\ell)$ -valued process with the following dynamics:*

- (i) *Blocks with the same label x coalesce to create a new block with that same label according to the (non-spatial) Λ_x -coalescent independently from blocks at other sites, meaning that coalescence happens with the rates given in (1) for $\Lambda = \Lambda_x$.*
- (ii) *The label of each block performs an independent random walk on G with the migration rate from site x to site y given by $p(x, y)$, $x, y \in G$.*

It is shown in Limic and Sturm (2006) Theorem 1 that the spatial Λ -coalescent is a well-defined strong Markov process for any initial condition in \mathcal{P}^ℓ in the case that Λ_x are the same measures for all $x \in G$ and $|G| < \infty$. However, the construction can easily be extended to this more general setting and also to $|G|$ countably infinite (some care must be taken here for starting configuration with infinitely many blocks, but we will not need this here). As mentioned earlier, the special case for which Λ_x is given by δ_0 up to a constant for all $x \in G$ is called the structured coalescent (or spatial Kingman coalescent).

3. Spatial Cannings models and derivation of spatial Λ coalescents

In this section we want to present a class of simple spatial population model, which we will refer to as *spatial Cannings models*. We will show that the spatial genealogy of a sample of individuals from various sites is described by the spatial Λ -coalescent of Definition 2.1 in the large population limit. Here, the number $|G|$ of sites in the graph may be finite or countably infinite. The convergence result combines work by Herbots (1997), who considered convergence to the structured coalescent, with results on convergence of genealogies to Λ -coalescents in the non-spatial setting from Möhle and Sagitov (2001).

In the spatial Cannings model there are $N_x := r_x N \in \mathbb{N}$ individuals at site $x \in G$ at any time, where r_x is a constant. These populations reproduce within their own site at discrete times $s \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$ and immediately after birth disperse their offspring to other sites. Both the reproduction as well as the dispersal or migration mechanism are chosen such that they leave the population size at any site fixed. The offspring law at site $x \in G$ is described by

$$\nu^x = (\nu_1^x, \nu_2^x, \dots, \nu_{N_x}^x), \quad (5)$$

where ν_l^x is the number of offspring of individual l in the previous generation (all at site x). In order to keep the population sizes constant over time we assume that $\sum_{l=1}^{N_x} \nu_l^x = N_x$.

For the offspring distributions ν^x we will furthermore assume that they are exchangeable such that for any permutation σ of the N_x indices we have

$$(\nu_1^x, \nu_2^x, \dots, \nu_{N_x}^x) \stackrel{\mathcal{D}}{=} (\nu_{\sigma(1)}^x, \nu_{\sigma(2)}^x, \dots, \nu_{\sigma(N_x)}^x). \quad (6)$$

Cannings (1974) first considered non-spatial versions of this model with offspring distributions satisfying these properties.

Due to the exchangeability and the criticality of the reproduction we have $\mathbb{E}(\nu_l^x) = 1$. The probability that any two individuals at site x have a common ancestor in the previous generation is given by

$$c_x^N := \sum_{l=1}^{N_x} \mathbb{E} \left(\frac{\nu_l^x(\nu_l^x - 1)}{N_x(N_x - 1)} \right) = \frac{\mathbb{E}((\nu_1^x)_2)}{N_x - 1} = \frac{\text{Var}(\nu_1^x)}{N_x - 1}, \quad (7)$$

where we are denoting $(m)_k := \frac{m!}{(m-k)!}$. After reproduction a fixed number of offspring n_{xy} selected at random without replacement from site x migrate to site y for all $x, y \in G$. Thus, for each $x \in G$, we need that $\sum_{y \in G} n_{xy} \leq N_x$. In order to keep the population sizes at all sites fixed even after the dispersal due to migration we also require *balancing migration*, meaning that

$$\sum_{y \neq x} n_{xy} = \sum_{y \neq x} n_{yx}. \quad (8)$$

Taking reproduction and migration together specifies the spatial Cannings model fully. We note that the special case considered by Herbots (1997) corresponds to Wright-Fisher reproduction in all colonies (which are taken to be of the same size N) such that ν_x of (5) is given by a symmetric multinomial distribution for all x .

In order to study the associated genealogies of a sample we still need the following notation. After reproduction and migration, the proportion of individuals in site x who were born in site y is given by

$$p_{xy} = \frac{n_{yx}}{N_x} = \frac{N_y q_{yx}}{N_x} = \frac{r_y}{r_x} q_{yx}, \quad (9)$$

where q_{yx} is the proportion of individuals in site y leaving for site x . We also set $p_x = \sum_{y \neq x} p_{xy}$, which is the proportion of individuals at site x that migrated there.

Now let the *spatial Cannings-coalescent* $\Pi^{\ell, N} = (\Pi_s^{\ell, N})_{s \in \mathbb{N}_0}$ be derived from a spatial Cannings model such that this \mathcal{P}_n^ℓ valued process is obtained from sampling n individuals at the present, whose location is represented by $\pi^\ell \in \mathcal{P}_n^\ell$, and then following their genealogy into the past.

In order to obtain a spatial Λ -coalescent in continuous time from Definition 2.1 in the large population limit as $N \rightarrow \infty$ we will rescale time for this process as a function of N and let $N \rightarrow \infty$. Set

$$t^N = \frac{t}{5 c^N}, \quad (10)$$

where c^N is related to c_x^N and specified later on. Naturally, if c_x^N are independent of $x \in G$ we will set $c^N = c_x^N$. This would in particular be the case in a spatially homogeneous situation in which all colonies are of the same size and display the same reproductive behavior, meaning that $r_x = 1$ so that $N_x = N$ and $\nu^N = \nu^{x,N}$ for all sites $x \in G$.

Further, we will need an asymptotic moment condition as in Möhle and Sagitov (2001) stating that

$$\phi_j^x(k_1, \dots, k_j) := \lim_{N \rightarrow \infty} (N^{k_1 + \dots + k_j - j} c_x^N)^{-1} \mathbb{E}((\nu_1^{x,N})_{k_1} \dots (\nu_j^{x,N})_{k_j}) \quad (11)$$

exist for all $x \in G$, $j \in \mathbb{N}$ and $k_1, \dots, k_j \geq 2$. We are now ready to state a convergence in distribution result for the spatial Cannings-coalescent.

Proposition 3.1. *Assume that there exists a sequence $\{c^N\}_{N \in \mathbb{N}} \subset \mathbb{R}_+$ with $\lim_{N \rightarrow \infty} c^N = 0$ and constants c_x such that for any $x \in G$ we have $\lim_{N \rightarrow \infty} \frac{c_x^N}{c^N} = c_x$ with $\sup_{x \in G} c_x < \infty$. Also assume that $\lim_{N \rightarrow \infty} \frac{p_{xy}^N}{c^N} = p(x, y)$ where $\sup_{x \in G} \sum_{y \neq x} p(x, y) < \infty$. Assume further that $\phi_1^x(k)$ exist for all $x \in G$ with $\sup_{x \in G} \phi_1^x(k) < \infty$ for all $k \geq 2$ as well as*

$$\phi_2^x(2, 2) = \lim_{N \rightarrow \infty} (N^2 c_x^N)^{-1} \mathbb{E}((\nu_1^{x,N})_2 \cdot (\nu_2^{x,N})_2) = 0 \quad (12)$$

for all $x \in G$. For $n \in \mathbb{N}$ and $\pi^\ell \in \mathcal{P}_n^\ell$ let $\Pi^{\ell,N}$ be the spatial Cannings-coalescent started in π^ℓ . Then we obtain weak convergence

$$(\Pi_{[\frac{t}{c^N}]}^{\ell,N})_{t \in \mathbb{R}_+} \Rightarrow (\Pi_t^\ell)_{t \in \mathbb{R}_+} \quad \text{in } D(\mathbb{R}_+, \mathcal{P}_n^\ell) \quad (13)$$

as $N \rightarrow \infty$, where Π^ℓ is the spatial Λ -coalescent as in Definition 2.1 also started in π^ℓ with Λ_x characterized by the moments given by

$$\int_0^1 z^{k-2} \Lambda_x(dz) = c_x \phi_1^x(k), \quad k \geq 2. \quad (14)$$

The condition (12) implies that $\phi_j^x \equiv 0$ for $j \geq 2$ and all $x \in G$ by monotonicity in j , see (18) of Möhle and Sagitov (2001). This also implies immediately that we obtain convergence to the spatial Kingman coalescent if and only if

$$\phi_1^x(3) = \lim_{N \rightarrow \infty} (N^2 c_x^N)^{-1} E((\nu_1^{N,x})_3) = 0. \quad (15)$$

As mentioned earlier, a special case of Proposition 3.1 was proved by Herbots (1997). For the Wright-Fisher reproduction she considers we set $c^N = \frac{1}{N}$. Then under the same assumptions Herbots shows convergence of the number of blocks on the space $D(\mathbb{R}_+, \mathbb{N}^{|G|})$. Since (15) is satisfied, the limit is the block counting process of the structured coalescent. Here, we consider convergence of spatial partition valued processes in a more general setting.

Proposition 3.1 in the non-spatial situation for which $|G| = 1$ was proved by Möhle and Sagitov (2001). They considered convergence of partition valued processes for even more general offspring distributions and proved that the limiting process Π is a coalescent with simultaneous multiple collisions, or Ξ -coalescent, if and only if (11) is satisfied. Proposition 3.1 can easily be extended to those settings as well but since we will, for ease of notation, only be dealing with spatial Λ -coalescents later on we restrict our attention to this subclass.

4. Behavior of the spatial Λ -coalescent on a large 2-dimensional torus

From now on let $G = T^L$ and in order to record the dependence on L we will write $\mathcal{P}_n^{\ell,L}$ and $\mathcal{P}^{\ell,L}$ for the partitions with labels in T^L . We will for notational simplicity assume that $\Lambda_x = \Lambda$ is constant for all $x \in T^L$ and that $\Lambda([0, 1]) > 0$. (It really suffices to assume that $\lambda_{2,2}^x$ is bounded above and below by positive constants uniformly in $x \in T^L$.) In order to specify the migration on the torus we first consider transition rates $\tilde{p} : \mathbb{Z}^2 \rightarrow [0, 1]$ and define for all $L \in \mathbb{N}$ transition rates p^L of a random walk on T^L by

$$p^L(x, y) = \sum_{z \in \{z' \in \mathbb{Z}^2 \mid z' - y \bmod 2L+1 = 0\}} \tilde{p}(x, z). \quad (16)$$

This means that ancestral lines that "migrate out" on one side of T^L will "migrate in" again on the other side. To simplify the arguments in the following we will assume that the rates \tilde{p} are those of a simple symmetric random walk on \mathbb{Z}^2 that migrates to each of the four nearest neighbour sites with equal rates $\frac{1}{4}$. However, any spatially homogeneous, symmetric random walk satisfying a suitable moment condition could also be considered (see also the remark at the end of this section).

On T^L we also define the following metric r_L appropriate for the torus,

$$r_L(x, y) := \inf\{\|x - z\|_2 \mid z \in \mathbb{Z}^2, y - z \bmod 2L+1 = 0\}, \quad (17)$$

where $\|x\|_2 := \sqrt{x_1^2 + x_2^2}$ for $x \in \mathbb{Z}^2$. Furthermore, we define for $0 \leq a \leq b$ the set of all labeled partitions with pairwise distances in $[a, b]$:

$$[[a, b]] := \{\pi^\ell = ((B_1, \zeta_1), (B_2, \zeta_2), \dots) \in \mathcal{P}_n^{\ell,L} \mid r_L(\zeta_i, \zeta_j) \in [a, b] \text{ for all } i \neq j \text{ with } B_i, B_j \neq \emptyset\}. \quad (18)$$

In order to describe the asymptotic behavior of the spatial Λ - n -coalescent started with n individuals we need some more notation: For $\pi \in \mathcal{P}_n$ let

$$K^\pi = (K_t^\pi)_{t \in \mathbb{R}_+} \quad (19)$$

be the *non-spatial* Kingman coalescent on $[n]$ started in π . We also define the sequence $(s_L)_{L \in \mathbb{N}}$, which will be used to rescale time, by

$$s_L := (2L+1)^2 \log(2L+1). \quad (20)$$

Our main theorem then states the following.

Theorem 4.1. *Let $n \geq 2$ and let $(a_L)_{L \in \mathbb{N}}$ be a nonnegative sequence with*

$$\lim_{L \rightarrow \infty} L^{-1} \sqrt{\log L} a_L = \infty, \quad \lim_{L \rightarrow \infty} L^{-1} a_L = 0$$

and $a_L \leq \sqrt{2}L$ for all $L \in \mathbb{N}$. Also, let $\pi_0 \in \mathcal{P}_n$ and $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ such that

$$\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$$

and $\pi^L = \pi_0$ for all $L \in \mathbb{N}$ large enough. For $L \in \mathbb{N}$ let $\Pi^{\ell,L}$ be a spatial Λ - n -coalescent started in $\pi^{\ell,L}$. Then, we obtain weak convergence in the Skorohod space $D(\mathbb{R}_+, \mathcal{P}_n)$ for $L \rightarrow \infty$, more precisely

$$(\Pi_{s_L t}^L)_{t \in \mathbb{R}_+} \Rightarrow (K_{\pi t}^{\pi_0})_{t \in \mathbb{R}_+}. \quad (21)$$

Note that we will generally assume that L is large enough so that $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ corresponding to a given $\pi_0 \in \mathcal{P}_n$ can be found. Observe also that in the limiting non-spatial Kingman coalescent there is an additional time change by the number π . The time change and so the entire limit process does not depend on the details of the coalescence mechanism of the spatial Λ -coalescent.

Finally, we remark that the result of Theorem 4.1 is also expected to hold if the random walk \tilde{p} on \mathbb{Z}^2 is symmetric and spatially homogeneous, and satisfies a suitable moment condition. In this case, the time scaling by π in (21) would have to be replaced by $\sigma^2 \pi$ with σ^2 the variance of distance traversed by the random walk in one unit of time.

5. Discussion

Knowledge of the genealogies together with an independent mutation process, in our context generally a Poisson process along the ancestral lineages, allows to describe the distribution of quantities of interest in population genetics that can be read off the genetic variability in the sample, such as site or allele frequency spectra.

In the non-spatial situation many results are available for the Kingman coalescent case modeling populations with low offspring variance, a prominent example being the Ewens sampling formula for the allele frequency spectrum, see for example Wakeley (2009) for an overview. For non-spatial Λ -coalescents modeling populations with larger offspring variances the analysis is more complicated and the available results are not as complete or explicit. However, see for example Möhle (2006a; 2006b), Berestycki et al. (2007), Birkner et al. (2011), and Berestycki et al. (unpublished manuscript) for a variety of recent results that show that differences in the underlying reproduction can lead to qualitatively different sampling distributions.

Likewise, the analysis of genetic variability in the sample becomes much more difficult due to spatial structure. Generally, the influence of the spatial structure will again lead to qualitatively different sampling distributions. However, this depends on the underlying space and in particular on the distances of the sampled individuals.

Our main results state that the genealogy of individuals sampled far enough apart on a large two dimensional torus T^L can be approximated by the genealogy of a *non-spatial* Kingman coalescent even when the offspring variances in the underlying population models are larger. On one hand, this means that the results available for the non-spatial Kingman coalescent can be used to approximate the genealogy and the resulting sampling distributions. However, from the point of view of a population geneticist it is a negative result: If one samples individuals relatively far apart then the influence of a large variance in the offspring distribution as well as of spatial structure are harder to detect in the sample.

In order to derive this result we consider spatial Λ -coalescents on T^L , which model the genealogies of individuals sampled from large populations living and migrating on T^L whose offspring distributions have potentially larger variances. In Section 3 we introduced spatial Cannings models as a large class of models that fit into this framework, a statement that is made precise by Proposition 3.1.

The main result, Theorem 4.1, then states that the (unlabeled) partition structure of the suitably time changed spatial Λ - n -coalescents on T^L converges to that of a non-spatial Kingman coalescent as $L \rightarrow \infty$ provided that the individuals are sampled far enough apart. This kind of behavior arises since in the chosen scaling and due to the sparse sample it is unlikely that more than two ancestral lines, represented by blocks, ever meet at the same site. Thus, only binary mergers may take place. On the other hand, a meeting of two ancestral lines is followed up by many more

meetings of these two so that they eventually coalesce (regardless of the rate of coalescence) before encountering any other ancestral lines. The sequence of meetings of two ancestral lines and thus also their coalescence happens instantaneously as $L \rightarrow \infty$, which implies that for large L the spatial Λ - n -coalescent behaves like a coalescing random walk (with instantaneous coalescence of lines that have met). In contrast, the time between encounters and eventual coalescence of pairs of lines is long enough so that all ancestral lines have in the meantime become well mixed on the torus. Hence, any two of them are equally likely to participate in the next meeting, which leads in the limit to the exchangeability property of the non-spatial Kingman coalescent.

The result of Theorem 4.1 is analogous to one obtained by Limic and Sturm (2006) for the d dimensional torus with $d \geq 3$. However, the scaling for $d = 2$ is different and more subtle. This is due to the fact that the random walk performed by the ancestral lines is recurrent in $d = 2$ while it is transient for $d \geq 3$. The recurrence in two dimensions also leads to the many encounters and in the limit $L \rightarrow \infty$ to instantaneous coalescence of a pair of lines that have met at the same site. As a consequence the time change of the limiting Kingman coalescent is independent of the Λ -measure of the underlying coalescent mechanism in $d = 2$ unlike in $d = 3$. The prior work by Limic and Sturm (2006) generalised results by Greven et al. (2005) who considered spatial Kingman coalescents in $d \geq 3$ (for some results in $d = 2$ in the analogous setting but with a different focus see also Greven et al. (2012)).

The observation that the influence of spatial structure on the genealogy of a sample is not readily detectable in certain situations (except possibly through a space dependent time change) has been made before. Related results have in particular been obtained in the articles by Cox and Durrett (2002) and Zähle et al. (2005). They study the classical stepping stone model on the torus T^L . This is a spatial Moran model and thus a special case of the spatial Cannings model introduced in Section 3 corresponding to setting $N_x = N$ and choosing $\nu_x^N = \nu^N$ in (5) as a permutation of $(2, 0, 1, \dots, 1)$. In Cox and Durrett (2002) and Zähle et al. (2005) meeting times and coalescence times of ancestral lineages are analysed as L tends to infinity while the populations size N and migration rate depend on L in an appropriate way. In this setting they also prove a result analogous to Theorem 4.1, stating that the genealogy of individuals sampled far enough apart can be approximated by a non-spatial Kingman coalescent.

A related model in continuous space has recently been introduced by Barton et al. (2010). Here, reproduction events are determined by Poisson point processes in space and involve individuals in a certain neighbourhood. Rare extinction-recolonisation events lead to a genealogy that is described by a spatial coalescent with a Λ -coalescent mechanism affecting ancestral lines in the neighbourhood chosen by the Poisson point process.

These Λ -coalescents in continuous space differ from the ones considered here in discrete space. Nevertheless, an analogous result to Theorem 4.1 is obtained when individuals sampled far enough apart on a two dimensional continuous torus are considered. As its side length L tends to infinity, the limiting genealogy is described by a non-spatial Kingman coalescent, another type of a spatial Λ -coalescent or coalescing Brownian motions with non-local coalescence, depending on the scaling of the neighbourhood sizes that affect multiple merger events.

In this work, the behavior of the spatial Λ -coalescent when individuals are not sampled far apart has not been considered analytically. It is clear, however, that due to the recurrence of the random walk in $d = 2$ coalescence of all ancestral lines would happen instantaneously on the time scale considered in Theorem 4.1 in the limit as $L \rightarrow \infty$. This is in contrast to results in $d \geq 3$ as in Limic and Sturm (2006), where there is a nontrivial distribution of ancestral lines (far apart from each other) for all large L .

For ancestral lines sampled close to each other the behavior of the ancestral lines is often

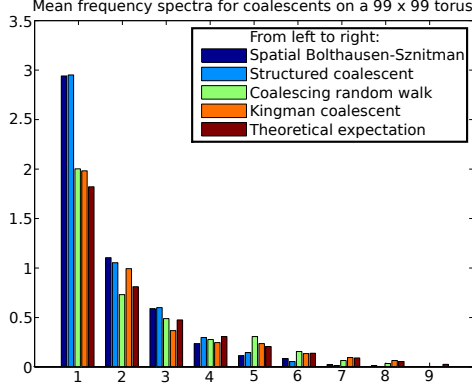


Figure 1: Mean frequency spectrum for $L' = 99$

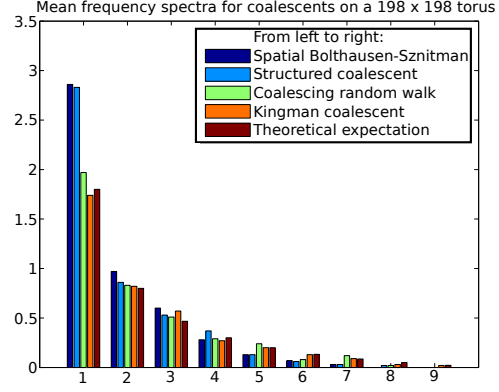


Figure 2: Mean frequency spectrum for $L' = 198$

described by distinguishing two phases, the first phase termed the *scattering phase* that (likely) involves initial rapid coalescence and lasts until ancestral lines are far apart (and well mixed) in space, and the second phase of subsequent slow coalescence termed the *collecting phase*, see Wakeley (2001). Given this terminology, our analytical results are restricted to the collecting phase. For some analytical results on the genealogy for the scattering and collecting phase under various model assumptions, see Zähle et al. (2005), Etheridge and Véber (in press 2012), and Greven et al. (unpublished manuscript).

In order to further examine the asymptotic results for the spatial Λ -coalescent we performed simulations of the spatial Λ -coalescent as well as the (non-spatial) Kingman coalescent on a moderately large two dimensional torus. We then compared the mean allele frequency spectrum of both processes which we juxtaposed additionally with its expectation in the Kingman coalescent calculated using Ewens sampling formula. We also compared the total tree length of both processes via a q-q-plot.

The allele frequency spectrum is computed under the infinite alleles model. In order to test how well the approximation performs for moderately large L mutations, that are assumed to always generate novel alleles, are placed on the genealogical tree of the processes according to a Poisson point process with the rate $\pi \cdot ((2L + 1)^2 \log(2L + 1))^{-1}$. We start with n singletons at time 0. Whenever a block is hit by a mutation we count the number of individuals in the block. Thus, if there are k individuals in a block that is hit by a mutation a k -tuple of individuals carrying this mutation is generated. We then remove the individuals from the system. When all individuals are removed we count the number of k -tuples, this number shall be called a_k . The vector (a_1, \dots, a_n) is the desired frequency spectrum. We perform m independent simulations and generate a_i^j , which represent the i -th component of the allele frequency spectrum of the j -th simulation. The mean frequency spectrum is now given by $m^{-1} \sum_{j=1}^m (a_1^j, \dots, a_n^j)$.

First, we start with $n = 9$ individuals sampled far apart on the torus at time 0. We arrange them on the torus in a 3×3 square such that the distance of next neighbours is a third of the side length $L' = 2L + 1$.

Regarding the various coalescence behaviors we consider the (instantaneously) coalescing random walk, the spatial Bolthausen-Sznitman coalescent (Λ uniform on $[0, 1]$) and the structured coalescent (the spatial Kingman coalescent with $\Lambda = \delta_0$). For the side lengths $L' = 99$ and

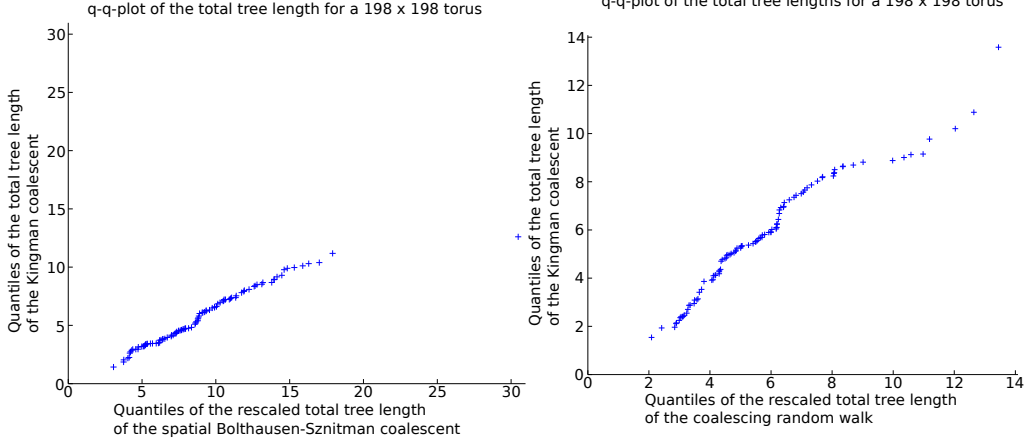


Figure 3: q-q-plot of the rescaled total tree length of the spatial Bolthausen-Sznitman coalescent versus the total tree length of the Kingman coalescent for $L' = 198$

Figure 4: q-q-plot of the rescaled total tree length of the coalescing random walk versus the total tree length of the Kingman coalescent for $L' = 198$

$L' = 198$ we perform $m = 100$ independent simulations.

Figures 1 and 2 show that the results for the coalescing random walk are much closer to the limiting non-spatial Kingman coalescent than those for the spatial Λ -coalescents. The spatial Λ -coalescents produce more singletons in the allele frequency spectrum in comparison with the Kingman coalescent. Moreover, they also produce a larger total tree length (compare figures 3 and 4).

Comparing figures 1 and 2 we note in addition that there is not much of an improvement between the $L' = 99$ and $L' = 198$ case, which is a sign for a slow convergence rate in Theorem 4.1. This can be explained by recalling the proof of the result one more time. We show the convergence result first for the coalescing random walk. We then show that for large L the spatial Λ -coalescent behaves like a coalescing random walk since any pair of ancestral lines that meet in the spatial Λ -coalescent will with high probability on a large torus meet again and again (before meeting other ancestral lines) until the pair eventually coalesces. This additional time until coalescence vanishes in the limit. Nonetheless, for moderately large L it is not surprising that we see longer coalescence times for the spatial Λ -coalescents than for the coalescing random walks and therefore also more singletons in the allele frequency spectrum since singletons have more time to get hit by a mutation before coalescing. Note that the behavior of the Bolthausen-Sznitman coalescent and the structured coalescent are quite similar since it is mostly $\lambda_{2,2}$ that dictates how much longer it takes the spatial Λ -coalescent for coalescence, and we have $\lambda_{2,2} = 1$ for both of these choices.

The slow convergence rate may be explained by looking at the expectation of the time that it takes for two blocks to leave the same site and then meet again. This expected time is of order L^2 and after rescaling it is of order $(\log L)^{-1}$. Therefore, these expectations converge only very slowly to zero.

Next, we consider individuals sampled close to each other on the torus and their behavior under the spatial Bolthausen-Sznitman coalescent. Once the ancestral lines are mutually far apart again, we also use the approximation by a non-spatial Kingman coalescent. Therefore we get two

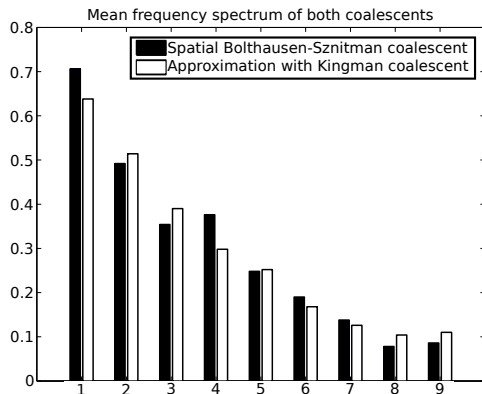


Figure 5: Mean frequency spectrum of the Bolthausen-Sznitman coalescent and the Approximation with the Kingman coalescent with ancestral lines started close to each other for $L' = 99$

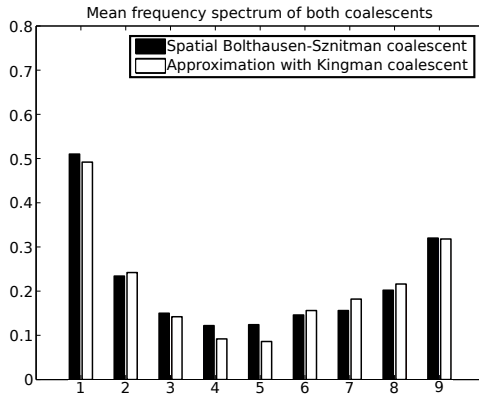


Figure 6: Mean frequency spectrum of the Bolthausen-Sznitman coalescent and the Approximation with the Kingman coalescent with ancestral lines started at the same position for $L' = 99$

results for the allele frequency spectrum, one given by the simulation of the spatial process and the other given by simulating until ancestral lines are far apart and then approximating with the Kingman coalescent.

Again we have $n = 9$ individuals to start with, the side length of the torus is $L' = 99$ and we repeat the simulation $m = 500$ times. We start the approximation when the ancestral lines surpass a mutual distance of 8.33. We either start the particles in a 3×3 scheme where next neighbours have distance 1 or we start them all in the same site.

For both cases we compute the mean frequency spectrum (see figures 5 and 6). In comparison with the sparse situation (figure 1) we observe that the mean of a_1 and a_2 decreases and the mean of a_6, a_7, a_8 and a_9 increases. This effect is expected and is even stronger for individuals started at the same site.

We observe again that the approximation with the Kingman coalescent gives less weight to singletons but since the number of ancestral lines left in the system when it has become sparse will be very small (possibly 1) the approximation is sometimes not even used so that the results appear to be better than in the case where we already start far apart. Here, the mean number of ancestral lines at the time of approximation was 2.6 for the close starting configuration and 4.0 for all ancestral lines started in the same site.

Finally, we note that if ancestral lines end up far apart the simulation of the spatial Λ -coalescent will take significantly more time than the simulation of the approximation with the non-spatial Kingman coalescent, and thus could be of practical use. Here, the simulation with the approximation was more than 100 times faster than the simulation without the approximation.

In conclusion, we recall that our theoretical convergence result states that information about spatial structure and the specific coalescence mechanism (due to possibly larger variances in the offspring distribution) is lost from a sparse sample as $L \rightarrow \infty$, except for a time change that depends on L . However, we have seen from simulations that even if this time change is not taken into account the influence of the spatial structure is still quite evident for moderately large L . In contrast, the details of the coalescence mechanism -apart from the delay that it takes a pair of lineages to coalesce while at the same site- do not have a significant impact on the sampling

distribution for moderately large L in the case of a sparse sample.

In future theoretical work, it would be interesting to study the rate of convergence, and also whether it can be improved by a finer scaling that takes the coalescence delay into account. Also, theoretical investigation of the scattering phase for non sparse samples as well as the consideration of convergence as N and L tend to infinity jointly would be of interest.

6. Proofs

In this section we prove Theorem 3.1 and Theorem 4.1.

6.1. Proof of Theorem 3.1

For the proof we fix the following notation. For a vector $a = (a_x)_{x \in G}$ we set $\|a\| = \sup_{x \in G} |a_x|$. For $(a^N)_{N \in \mathbb{N}}$ and $(c^N)_{N \in \mathbb{N}}$ sequences of vectors and nonnegative numbers respectively we say that $a^N = O(c^N)$ if $\sup_{N \in \mathbb{N}} \frac{\|a^N\|}{c^N} < \infty$, and $a^N = o(c^N)$ if $\lim_{N \rightarrow \infty} \frac{\|a^N\|}{c^N} = 0$. If $A = (a_{ij})_{i,j \in I}$ is a matrix for a countable I then we define the matrix norm by

$$\|A\| = \sup_{i \in I} \sum_{j \in I} |a_{ij}|. \quad (22)$$

In this proof we only consider labeled partitions and thus omit the superscript ℓ in the notation whenever possible. So for $n \leq N$ let $\pi, \tilde{\pi} \in \mathcal{P}_n^\ell$ and let $p_{\pi\tilde{\pi}}^N = \mathbb{P}(\Pi_{t+1}^{\ell,N} = \tilde{\pi} | \Pi_t^{\ell,N} = \pi)$ be the transition probabilities of $\Pi^{\ell,N}$. We set $P_N = (p_{\pi\tilde{\pi}}^N)_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$. Also let $q_{\pi\tilde{\pi}}$ be the transition rate from π to $\tilde{\pi}$ for Π^ℓ if $\pi \neq \tilde{\pi}$ and set $q_{\pi\pi} = -\sum_{\tilde{\pi} \neq \pi} q_{\pi\tilde{\pi}}$ as well as $Q = (q_{\pi\tilde{\pi}})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$. For the proof of the convergence of finite dimensional distributions it suffices to show that

$$P_N = I + c^N Q + o(c^N). \quad (23)$$

The sufficiency follows since $\|P_N\| = 1$ and also $\|I + c^N Q\| = 1$ for N large due to the fact that the assumptions of the proposition imply that $q_{\pi\tilde{\pi}}$ are uniformly bounded over all $\pi, \tilde{\pi} \in \mathcal{P}_n^\ell$. From this and (23) we then have

$$\|P_N^{[\frac{t}{c^N}]} - (I + c^N Q)^{[\frac{t}{c^N}]}\| \leq \frac{t}{c^N} \|P_N - (I + c^N Q)\| = t \left\| \frac{P_N - I}{c^N} - Q \right\| \rightarrow 0 \quad (24)$$

as $N \rightarrow \infty$ which implies that

$$\lim_{N \rightarrow \infty} P_N^{[\frac{t}{c^N}]} = \lim_{N \rightarrow \infty} (I + c^N Q)^{[\frac{t}{c^N}]} = e^{tQ}. \quad (25)$$

This in turn implies convergence of the finite dimensional distributions.

Thus, we will first focus on showing relation (23). For notational simplicity we will in the following omit a sub- or superscript N whenever this does not affect clarity such as for the quantities p_{xy}^N and p_x^N . We can write $P_N = P_N^{(m)} \cdot P_N^{(r)}$ where $P_N^{(m)} = (p_{\pi\tilde{\pi}}^{(N,m)})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$ and $P_N^{(r)} = (p_{\pi\tilde{\pi}}^{(N,r)})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$ are the transition probability matrices of the coalescent due to the migration step and reproduction step respectively. We first consider the migration step. The arguments will be similar to those in the proof of Theorem 2.1 in Herbots (1997). If $\tilde{\pi} \in A_{xy}^\pi$ the set of all partitions that result from π by changing one label from x to y with $x \neq y$, then we have that

$$p_{\pi\tilde{\pi}}^{(N,m)} = \frac{\binom{N_x - \#\pi_x}{p_{xy}N_x - 1}}{\binom{N_x}{p_{xy}N_x}} \cdot \frac{\binom{N_x - p_{xy}N_x - \#\pi_x + 1}{p_xN_x - p_{xy}N_x}}{\binom{N_x - p_{xy}N_x}{p_xN_x - p_{xy}N_x}} \prod_{z \neq x} \frac{\binom{N_z - \#\pi_z}{p_zN_z}}{\binom{N_k}{p_zN_k}}. \quad (26)$$

Here, the first factor is the probability of choosing one particular block (but no other blocks) as migrants from site y to site x forward in time. The second factor represents the probability that all other migrants from other sites to x are drawn from outside the sample. Lastly, the product represents the probability that no migrants at sites other than x are in the sample. We can simplify this expression to

$$p_{\pi\tilde{\pi}}^{(N,m)} = p_{xy} \frac{N_x}{N_x - p_x N_x - \#\pi_x + 1} \prod_{z \in G} \prod_{a=0}^{\#\pi_z-1} \frac{N_z - p_z N_z - a}{N_z - a}. \quad (27)$$

We also have that

$$\sum_{\tilde{\pi} \notin \bigcup_{x \neq y} A_{xy}^\pi \cup \{\pi\}} p_{\pi\tilde{\pi}}^{(N,m)} \leq R^{(N,m)}(\pi), \quad (28)$$

where $R^{(N,m)}(\pi)$ is the probability that at least two of the migrants are drawn from the sample. As in Herbots (1997) we can bound

$$R^{(N,m)}(\pi) \leq \left(\sum_{z \in G} \#\pi_z p_z \right)^2 \leq n^2 \sup_{z \in G} p_z^2 = O((c^N)^2) \quad (29)$$

by assumption. Taking (27) to (29) together now implies that

$$P_N^{(m)} = I + c^N Q_N^{(m)} + R_N^{(m)}, \quad (30)$$

where the entries of $Q_N^{(m)} = (q_{\pi\tilde{\pi}}^{(N,m)})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$ are given by $q_{\pi\tilde{\pi}}^{(N,m)} = \frac{1}{c^N} p_{\pi\tilde{\pi}}^{(N,m)}$ if $\tilde{\pi} \in A_{xy}^\pi$ for some $x \neq y$, by $q_{\pi\pi}^{(N,m)} = -\sum_{x \neq y} \sum_{\tilde{\pi} \in A_{xy}^\pi} \frac{1}{c^N} p_{\pi\tilde{\pi}}^{(N,m)}$ and are zero otherwise, and $R_N^{(m)}$ is the matrix containing the rest terms. We have due to (28) and (29) that $\|R_N^{(m)}\| = O((c^N)^2)$. Before turning to transition probabilities due to reproduction let us observe that

$$\lim_{N \rightarrow \infty} q_{\pi\tilde{\pi}}^{(N,m)} = q_{\pi\tilde{\pi}}^{(m)}, \quad (31)$$

where $q_{\pi\tilde{\pi}}^{(m)} = p(x, y)$ if $\tilde{\pi} \in A_{xy}^\pi$ and is zero for all other $\tilde{\pi} \neq \pi$. This convergence follows directly from (27) since the term multiplying p_{xy} in (27) is a finite product whose individual factors converge to 1 due to the fact that $N_z \rightarrow \infty$ and $p_z = O(c^N) \rightarrow 0$ by assumption for each $z \in G$. From (31) it then follows that $q_{\pi\pi}^{(m)} = -\sum_{\tilde{\pi} \neq \pi} q_{\pi\tilde{\pi}}^{(m)}$ so that $Q_N^{(m)} = (q_{\pi\tilde{\pi}}^{(m)})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$ is the matrix of the transition rates for the migration in the limit.

Recall now that $P_N^{(r)} = (p_{\pi\tilde{\pi}}^{(N,r)})_{\pi, \tilde{\pi} \in \mathcal{P}_n^\ell}$ describe the transition probabilities due to coalescence in the sample as a consequence of the Cannings reproduction forward in time. If $\tilde{\pi}$ arises from π by merging $2 \leq k = \#\pi - \#\tilde{\pi} + 1$ blocks with the same label x then (28) in Möhle and Sagitov (2001) states that

$$q_{\pi\tilde{\pi}}^{(r)} := \lim_{N \rightarrow \infty} \frac{1}{c^N} p_{\pi\tilde{\pi}}^{(N,r)} = \lim_{N \rightarrow \infty} \frac{c_x^N}{c^N} \cdot \frac{1}{c_x^N} p_{\pi\tilde{\pi}}^{(N,r)} = c_x \lambda_{\#\pi, k}^x. \quad (32)$$

with $\lambda_{b,k}^x$ defined as in (1) with Λ_x instead of Λ . For other $\tilde{\pi} \neq \pi$ we have that $q_{\pi\tilde{\pi}}^{(r)} := \lim_{N \rightarrow \infty} \frac{1}{c^N} p_{\pi\tilde{\pi}}^{(N,r)} = 0$. Setting $q_{\pi\pi}^{(r)} = -\sum_{\tilde{\pi} \neq \pi} q_{\pi\tilde{\pi}}^{(r)}$ we see that relation (32) implies the analogous statement to (30) for coalescence,

$$P_N^{(r)} = I + c^N Q_N^{(r)} + R_N^{(r)}, \quad (33)$$

where $R_N^{(r)} = (r_{\pi\tilde{\pi}}^{(N,r)})_{\pi,\tilde{\pi} \in \mathcal{P}_n^\ell}$ is the matrix of the rest terms. Since there are only finitely many (at most 2^n) non-zero entries in each row we obtain $\sum_{\tilde{\pi} \in \mathcal{P}_n^\ell} |r_{\pi\tilde{\pi}}^{(N,r)}| = o(c^N)$. Taken together this leads to $\|R_N^{(r)}\| = o(c^N)$. This implies that

$$P_N = I + c^N(Q_N + R_N), \quad (34)$$

where

$$Q_N = Q_N^{(m)} + Q^{(r)}, \quad (35)$$

$$R_N = \frac{1}{c^N}(R_N^{(m)} + R_N^{(r)} + R_N^{(m)}R_N^{(r)} + c^N Q_N^{(m)}R_N^{(r)} + c^N R_N^{(m)}Q^{(r)} + (c^N)^2 Q_N^{(m)}Q^{(r)}). \quad (36)$$

with $\|R_N\| \rightarrow 0$ as $N \rightarrow \infty$ since $\|R_N^{(m)}\| = O((c^N)^2)$ and $\|R_N^{(r)}\| = o(c^N)$. This and the convergence of $Q_N^{(m)}$ to $Q^{(m)}$ from (31) now completes the proof of (23) and so of convergence of finite dimensional distributions.

In order to complete the proof of Theorem 3.1 which states convergence in the Skorohod space $D(\mathbb{R}_+, \mathcal{P}_n^\ell)$, it remains to show relative compactness in $D(\mathbb{R}_+, \mathcal{P}_n^\ell)$. According to Corollary 3.7.4 in Ethier and Kurtz (1986) we need to show the following two conditions:

- (i) For every $\varepsilon > 0$ and $t \geq 0$ there exists a compact set $\Gamma_{\varepsilon,t} \subseteq \mathcal{P}_n^\ell$ such that

$$\liminf_{N \rightarrow \infty} \mathbb{P}(\Pi_{[\frac{t}{c^N}]}^{\ell,N} \in \Gamma_{\varepsilon,t}) \geq 1 - \varepsilon.$$

- (ii) For every $\varepsilon > 0$ and $t \geq 0$ there exists a $\delta > 0$ such that

$$\limsup_{N \rightarrow \infty} \mathbb{P}(w(\Pi_{[\frac{t}{c^N}]}^{\ell,N}, \delta, T) \geq \varepsilon) \leq \varepsilon$$

where

$$w(\Pi_{[\frac{t}{c^N}]}^{\ell,N}, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{s,t \in [t_{i-1}, t_i]} d_n(\Pi_{[\frac{s}{c^N}]}^{\ell,N}, \Pi_{[\frac{t}{c^N}]}^{\ell,N}), \quad (37)$$

and $\{t_i\}$ ranges over all partitions of the form $0 = t_0 < t_1 < \dots < t_{k-1} < T \leq t_k$ with $\min_{1 \leq i \leq k} (t_i - t_{i-1}) > \delta$ and $k \geq 1$.

Since \mathcal{P}_n^ℓ is finite, it is compact and condition (i) is trivially fulfilled. To prove condition (ii) we set $C := \|Q\|$ which is finite by assumption. Let $a_N = c_N(C + 1)$ and let in the following N be large enough so that $a_N < 1$ and $\|Q_N + R_N\| \leq C + 1$. Now consider the discrete time Markov chain $(E_t^N, e_t^N)_{t \in \mathbb{N}}$ which does not change its state with probability $1 - a_N$, changes only e_t^N to $e_{t+1}^N = e_t^N + 1$ with probability $a_N - \sum_{\tilde{\pi} \neq \pi} p_{\pi\tilde{\pi}}^{(N)} \geq 0$ if $E_{t+1}^N = E_t^N = \pi$. Finally, we have $e_{t+1}^N = e_{t+1}^N$ and $E_{t+1}^N = \tilde{\pi}$ given that $E_t^N = \pi$ with probability $p_{\pi\tilde{\pi}}^{(N)}$. This means that the process E^N is just a version of $\Pi^{\ell,N}$ if the time between the steps is chosen to be c_N . We have that the times between jumps of the Markov chain (E^N, e^N) are given by i.i.d geometric random variables τ_i^N with mean $\frac{1}{a_N}$. Since $w(e_{[\frac{t}{c^N}]}^N, \delta, T) = 0$ whenever there exists a J such that $\tau_i^N > \frac{\delta}{c^N}$ for $i = 1, \dots, J$ and $\sum_{i=1}^J \tau_i^N \geq \frac{T}{c^N}$ it suffices to show that we can find J and δ such that

$$\liminf_{N \rightarrow \infty} \mathbb{P}(\tau_i^N > \frac{\delta}{c^N} \text{ for } i = 1, \dots, J \text{ and } \sum_{i=1}^J \tau_i^N \geq \frac{T}{c^N}) \geq 1 - \eta.$$

This can be achieved since

$$\begin{aligned}
& \mathbb{P}(\tau_i^N > \frac{\delta}{c^N} \text{ for } i = 1, \dots, J \text{ and } \sum_{i=1}^J \tau_i^N \geq \frac{T}{c^N}) \\
&= \mathbb{P}(\sum_{i=1}^J \tau_i^N \geq \frac{T}{c^N} | \tau_i^N > \frac{\delta}{c^N} \text{ for } i = 1, \dots, J) \cdot \mathbb{P}(\tau_i^N > \frac{\delta}{c^N} \text{ for } i = 1, \dots, J) \\
&\geq \mathbb{P}(\sum_{i=1}^J \tau_i^N \geq \frac{T}{c^N}) \cdot \mathbb{P}(\tau_i^N > \frac{\delta}{c^N})^J \rightarrow \mathbb{P}(Z < J) \cdot \mathbb{P}(X > \delta)^J,
\end{aligned}$$

where Z is a Poisson random variable with mean $T(C + 1)$ and X is exponential with mean $\frac{1}{C+1}$. This finishes the proof. \square

6.2. Proof of Theorem 4.1

In the following let $\Pi^{\ell,L}$ be a spatial Λ - n -coalescent on T^L with transition kernel p^L . We will write \mathbb{P}^π to indicate that $\Pi^{\ell,L}$ is started in $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ at time 0. We want to show that asymptotically $\Pi^{\ell,L}$ behaves like a coalescing random walk for which only ever two blocks meet and coalesce instantaneously, and that the meetings only take place after the blocks have been "randomized" in space such that any two are equally likely to meet at any such event. In order to prove this we define another process of coalescing random walks $\tilde{\Pi}^{\ell,L}$ coupled to $\Pi^{\ell,L}$ (defined on the same probability space) and show that asymptotically $\Pi^{\ell,L}$ behaves like $\tilde{\Pi}^{\ell,L}$ and that $\tilde{\Pi}^{\ell,L}$ has the desired properties. Thus, the evolution of Π^L (unlabeled) is given by a non-spatial Kingman coalescent as $L \rightarrow \infty$.

We first couple the spatial Λ - n -coalescent with a coalescing random walk. Let $L \in \mathbb{N}$ and let $(\xi(i)^L)_{i \in [n]}$ be an i.i.d. family of random walks on T^L with transition kernel p^L . We now define a version of the spatial Λ - n -coalescent by stipulating that the label of a block B follows the random walk $\xi^L(\min(B))$.

We now require some notation. Let $i, j \in [n]$ and $t \in \mathbb{R}_+$. Let $A_t^L(i)$ define the unique block of $\Pi_t^{\ell,L}$ with $i \in A_t^L(i)$ and let $M_t^L(i) = \xi_t^L(\min(A_t^L(i)))$ be the location (label) of this block. We set

$$\tau^L(i, j) := \inf \{t \geq 0 \mid M_t^L(i) = M_t^L(j)\}, \quad (38)$$

$$\tau_c^L(i, j) := \inf \{t \geq 0 \mid A_t^L(i) = A_t^L(j)\}. \quad (39)$$

In words, these are the first times at which the two blocks containing i and containing j meet and coalesce, respectively. We also define the jump times

$$0 = \tau_0^L < \tau_1^L < \dots \quad \text{and} \quad 0 = \tau_{c,0}^L < \tau_{c,1}^L < \dots \quad (40)$$

where

$$\tau_{k+1}^L := \inf_{i,j \in [n]} \{ \tau^L(i, j) \mid \tau^L(i, j) > \tau_k^L \}$$

is the first meeting time after τ_k^L of blocks that have not met before and $\tau_{c,k+1}^L$ defined analogously is the first time after $\tau_{c,k}^L$ at which blocks coalesce.

Assuming that no two blocks of the starting configuration have the same label, we define the coalescing random walk $\tilde{\Pi}^{\ell,L}$ by first setting $\tilde{\Pi}_0^{\ell,L} = \Pi_0^{\ell,L}$. The label of a block \tilde{B} in $\tilde{\Pi}^{\ell,L}$ is following the random walk $\xi^L(\min(\tilde{B}))$ and as soon as blocks have the same label they coalesce

instantaneously. We set \tilde{M}^L and \tilde{A}^L as before, now for the process $\tilde{\Pi}^{\ell,L}$. We define the quantities in (38) to (40) analogously for the process $\tilde{\Pi}^{\ell,L}$ but note that $\tilde{\tau}^L(i, j) = \tilde{\tau}_c^L(i, j)$ and $\tilde{\tau}_k^L = \tilde{\tau}_{c,k}^L$. Furthermore note that $\tilde{\tau}_k^L = \tau_k^L$ for $k = 1$ but that the analogous statement is not true for $k > 1$ since blocks meeting in the spatial Λ - n -coalescent could part without coalescing. We will see later though that $\tilde{\tau}_k^L = \tau_k^L$ becomes very likely for large L if the blocks are initially sampled far apart.

We will now prove that the spatial Λ - n -coalescent behaves after rescaling like the coupled coalescing random walk. We define for any random walk Z^L on T^L with transition kernel p^L the stopping times

$$H^L := \inf \{t \geq 0 \mid Z_t^L = 0\}, \quad (41)$$

which is the *first hitting time of the origin* and

$$W^L := \inf \{t \geq 0 \mid Z_0^L = Z_t^L = 0 \text{ and there is an } s \in [0, t] \text{ with } Z_s^L \neq 0\}, \quad (42)$$

the *first return time to the origin*. We also use these definitions for $L = \infty$ with the convention that $T^\infty = \mathbb{Z}^2$.

Lemma 6.1. *Let $i, j \in [n]$. Then a.s.*

$$\tau_c^L(i, j) - \tau^L(i, j) = \tau_{\text{coal}} + \sum_{i=1}^{A-1} W_i^L, \quad (43)$$

where W_1^L, W_2^L, \dots are i.i.d copies of W^L of a random walk Z^L on T^L with transition kernel $2p^L$. The random variables τ_{coal} and A are independent from $\{W_i^L\}_{i \in \mathbb{N}}$ and from each other and $\tau_{\text{coal}} \sim \exp(\lambda_{2,2})$ as well as $A \sim \text{geom}(\lambda_{2,2}(2 + \lambda_{2,2})^{-1})$.

Proof. It is clear that $(M_t^L(i) - M_t^L(j))_{t \in \mathbb{R}_+}$, the distance between the blocks containing i and j , is a rate 2 random walk with transition kernel p^L until the coalescence time $\tau_c^L(i, j)$. Using the strong Markov property we restart this random walk at time $\tau^L(i, j)$, at which the blocks containing i and j meet for the first time. Therefore we restart in 0. The time until one block migrates away from the other block is exponentially distributed with parameter 2, the time until they coalesce while in the same site is exponentially distributed with parameter $\lambda_{2,2}$. Thus, the probability that the two coalesce before they part is given by $\lambda_{2,2}(2 + \lambda_{2,2})^{-1}$ and this then happens after a time τ_{coal} . If they part it will take time W_1^L to return to the same site. Using the strong Markov property we can repeat the argument leading to $A - 1$ independent return times W_i^L with A as above, and thus to the statement of the lemma. \square

As an immediate corollary we obtain that on the time scale that we are considering (recall the definition of s_L in (20)) the time from the first encounter of two blocks until their eventual coalescence is asymptotically short.

Corollary 6.2. *Let $i, j \in [n]$. For all $\varepsilon > 0$ and $\delta > 0$ there is an $L(\varepsilon, \delta) \in \mathbb{N}$, such that for all $L \geq L(\varepsilon, \delta)$ and for all $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ we have*

$$\mathbb{P}^{\pi^{\ell,L}}(s_L^{-1} |\tau_c^L(i, j) - \tau^L(i, j)| \geq \varepsilon) < \delta.$$

Proof. Let $L \in \mathbb{N}$ and $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$. If $A_0(i) = A_0(j)$ then $\tau_c^L(i, j) = 0 = \tau^L(i, j)$ and the claim follows. Now let $A_0(i) \neq A_0(j)$ and let $W_1^\infty, W_2^\infty, \dots$ be as in Lemma 6.1 for $L = \infty$. If we consider the random walk on T^L as the projection of a random walk on \mathbb{Z}^2 that defines the W_i^∞ then it follows that $W_i^L \leq W_i^\infty$. With Lemma 6.1 we have for all $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ that

$$\begin{aligned} \mathbb{P}^{\pi^{\ell,L}}(s_L^{-1} |\tau_c^L(i, j) - \tau^L(i, j)| \geq \varepsilon) &= \mathbb{P}^{\pi^{\ell,L}} \left(\tau_{\text{coal}} + \sum_{i=1}^{A-1} W_i^L \geq s_L \varepsilon \right) \\ &\leq \mathbb{P}^{\pi^{\ell,L}} \left(\tau_{\text{coal}} + \sum_{i=1}^{A-1} W_i^\infty \geq s_L \varepsilon \right). \end{aligned}$$

Since the random walk on \mathbb{Z}^2 is recurrent we have $W_i^\infty < \infty$ almost surely. Likewise τ_{coal} and A are almost surely finite. Since all these random variables do not depend on L and $\pi^{\ell,L}$ the claim follows now with $s_L \xrightarrow{L \rightarrow \infty} \infty$. \square

Let $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ and let every two blocks in $\pi^{\ell,L}$ have different labels. It is clear, that in the coalescing random walk $\tilde{\Pi}^{\ell,L}$ starting in $\pi^{\ell,L}$ almost surely only pairwise mergers happen. So for $k \in [\#\pi^{\ell,L} - 1]$ we have that $\tilde{\tau}_k^L < \infty$ is a time at which two blocks in $\tilde{\Pi}^{\ell,L}$ merge. By swapping the random walks belonging to the two blocks merged at time $\tilde{\tau}_k^L$ after time $\tilde{\tau}_k^L$ we can define a new coalescing random walk which agrees with $\tilde{\Pi}^{\ell,L}$ until time $\tilde{\tau}_k^L$ and which has the same distribution as $\tilde{\Pi}^{\ell,L}$. More precisely, this means that if \tilde{B}_1 and \tilde{B}_2 are the two blocks that meet at time $\tilde{\tau}_k^L$ with $\min \tilde{B}_1 < \min \tilde{B}_2$ then the label of the newly created block follows the motion $\xi_t^L(\min \tilde{B}_2)$ rather than $\xi_t^L(\min \tilde{B}_1)$. Let $\tilde{\tau}_{k+1}^{L,(2)}$ be the first meeting time of blocks in this process after time $\tilde{\tau}_k^L$. For $q > 0$ we consider the event

$$B^L := B^L(q, \pi^{\ell,L}) := \{\tilde{\tau}_{k+1}^L - \tilde{\tau}_k^L > s_L q, \tilde{\tau}_{k+1}^{L,(2)} - \tilde{\tau}_k^L > s_L q \text{ for all } k \in [\#\pi^{\ell,L} - 2]\}, \quad (44)$$

on which (after rescaling) it takes some time until a block that is created in a coalescence event meets further blocks whether it follows the motion of the first or the second block that took part in the coalescence event. Note that for $\#\pi^{\ell,L} = 2$ we have $B^L(q, \pi^{\ell,L}) = \Omega$ since no conditions need to be met.

Proposition 6.3. *Let $q > 0$. For all $\varepsilon > 0$, $\delta > 0$ and $\nu > 0$ there are constants $L(\varepsilon, \delta, \nu) \in \mathbb{N}$ so that for all $L \geq L(\varepsilon, \delta, \nu)$ and all $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ with $\#\pi^{\ell,L} \geq 2$, such that no two blocks of $\pi^{\ell,L}$ carry the same label, we have*

$$\mathbb{P}^{\pi^{\ell,L}} \left(s_L^{-1} \|(\tau_{c,1}^L, \dots, \tau_{c,n-1}^L) - (\tilde{\tau}_1^L, \dots, \tilde{\tau}_{n-1}^L)\| \geq \varepsilon, B^L(q, \pi^{\ell,L}) \right) < \delta \quad (45)$$

as well as

$$\mathbb{P}^{\pi^{\ell,L}} \left(\Pi_{\tau_{c,k}^L}^L \neq \tilde{\Pi}_{\tilde{\tau}_k^L}^L \text{ for some } k = 1, \dots, \#\pi^{\ell,L} - 1, B^L(q, \pi^{\ell,L}) \right) < \nu. \quad (46)$$

Note that the last statement (46) is about the unlabeled partition structures of the two processes. For proving this statement we set for $\Pi^{\ell,L}$ started in $\pi^{\ell,L}$

$$C^L(0) := \{(i, j) \in [n]^2 \mid i \sim_{\pi^{\ell,L}} j\} \quad (47)$$

the set of index pairs that are initially in different blocks

Proof. For any $q > 0$ there are almost surely exactly $\#\pi^{\ell,L} - 1$ coalescence events of $\tilde{\Pi}^{\ell,L}$ started in $\pi^{\ell,L}$ since at any coalescence event exactly two blocks merge. We will first show the following statements by induction in $k \in [\#\pi^{\ell,L} - 1]$:

There is an $L(\varepsilon, \delta, \nu) \in \mathbb{N}$ such that for all $\pi^{\ell,L}$ and $L \geq L(\varepsilon, \delta, \nu)$,

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_{c,k}^L - \tilde{\tau}_k^L \geq s_L \varepsilon, B^L(q, \pi^{\ell,L})) < \delta, \quad (48)$$

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_{c,k}^L < \tilde{\tau}_k^L, B^L(q, \pi^{\ell,L})) < \delta, \quad (49)$$

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_{c,k}^L \geq \tau_{k+1}^L, B^L(q, \pi^{\ell,L})) < \nu, \quad (50)$$

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_{k+1}^L \neq \tilde{\tau}_{k+1}^L, B^L(q, \pi^{\ell,L})) < \nu, \quad (51)$$

$$\mathbb{P}^{\pi^{\ell,L}}(\Pi_{\tau_{c,k}^L}^{\ell,L} \neq \tilde{\Pi}_{\tau_{c,k}^L}^{\ell,L}, B^L(q, \pi^{\ell,L})) < \nu. \quad (52)$$

These statements will then complete the proof since (45) follows immediately from (48) and (49) for all $k \in [\#\pi^{\ell,L} - 1]$. Likewise, (46) follows from (49), (50), and (52) for all $k \in [\#\pi^{\ell,L} - 1]$.

In order to start the induction let first $k = 1$. We begin by showing (48). Due to Corollary 6.2 we can find $L(\varepsilon, \delta) \in \mathbb{N}$ with

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_c^L(i, j) - \tau^L(i, j) \geq s_L \varepsilon \text{ for all } (i, j) \in [n]^2) < \frac{\delta}{n^2}$$

for all $L \geq L(\varepsilon, \delta)$ and all $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ with $\#\pi^{\ell,L} \geq 2$. Let $(i, j) \in C^L(0)$ so that $\tau_c^L(i, j) \geq \tau_{c,1}^L$. Thus, on the event $A_{i,j} := \{\tau_1^L = \tau^L(i, j)\}$, it follows that

$$\begin{aligned} \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tau_1^L \geq s_L \varepsilon, A_{i,j}, B^L(q, \pi^{\ell,L})) &= \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tau^L(i, j) \geq s_L \varepsilon, A_{i,j}, B^L(q, \pi^{\ell,L})) \\ &\leq \mathbb{P}^{\pi^{\ell,L}}(\tau_c^L(i, j) - \tau^L(i, j) \geq s_L \varepsilon) < \frac{\delta}{n^2}. \end{aligned}$$

Hence, with $\mathbb{P}^{\pi^{\ell,L}}(A_{i,j} \text{ for one } (i, j) \in C^L(0)) = 1$ we have

$$\mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tau_1^L \geq s_L \varepsilon, B^L(q, \pi^{\ell,L})) \leq \sum_{(i,j) \in [n]^2} \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tau_1^L \geq s_L \varepsilon, A_{i,j}, B^L(q, \pi^{\ell,L})) < \delta$$

for all $L \geq L(\varepsilon, \delta)$ and all $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ with $\#\pi^{\ell,L} \geq 2$, which is (48) since $\tilde{\tau}_1^L = \tau_1^L$.

Again because of $\tilde{\tau}_1^L = \tau_1^L$ statement (49) is trivial for $k = 1$ since $\tau_{c,1}^L \geq \tau_1^L$ by definition. For showing (50) assume that $\#\pi^{\ell,L} \geq 3$ as the statement is immediate for $\#\pi^{\ell,L} = 2$. Observe that we have $\tau_2^L = \tilde{\tau}_2^L$ or $\tau_2^L = \tilde{\tau}_2^{L,(2)}$. In both cases the arguments are completely analogous so it suffices to show the result for $\tilde{\tau}_2^L$ instead of τ_2^L , and hence to consider the event $\tau_{c,1}^L \geq \tilde{\tau}_2^L$.

We have $\tau_1^L = \tilde{\tau}_1^L < \tilde{\tau}_2^L < \infty$ almost surely. From (48) for $k = 1$ and $\varepsilon = q$ we obtain that there is an $L(\nu) = L(q, \nu) \in \mathbb{N}$ so that for all $L \geq L(\nu)$ and $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ with $\#\pi^{\ell,L} \geq 3$,

$$\begin{aligned} \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L \geq \tilde{\tau}_2^L, B^L(q, \pi^{\ell,L})) &= \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tilde{\tau}_1^L \geq \tilde{\tau}_2^L - \tilde{\tau}_1^L \geq s_L q, B^L(q, \pi^{\ell,L})) \\ &\leq \mathbb{P}^{\pi^{\ell,L}}(\tau_{c,1}^L - \tilde{\tau}_1^L \geq s_L q, B^L(q, \pi^{\ell,L})) < \nu, \end{aligned}$$

which proves (50).

For showing (51) and (52) assume again that $\#\pi^{\ell,L} \geq 3$ as the statement is trivially true for $\#\pi^{\ell,L} = 2$. We observe that from $\tau_{c,1}^L < \tau_2^L$ it follows that there is a coalescence event for $\Pi^{\ell,L}$

before any additional blocks meet. Thus, exactly the two blocks that met at time τ_1^L will have coalesced at time $\tau_{c,1}^L$ before meeting other blocks implying $\Pi_{\tau_{c,1}^L}^{\ell,L} = \tilde{\Pi}_{\tau_{c,1}^L}^{\ell,L}$ and $\tau_2^L = \tilde{\tau}_2^L$. Thus, (51) and (52) follow from (50).

Let the claim now be true for $k \in [\#\pi^{\ell,L} - 2]$ and let us consider $k + 1 \in [\#\pi^{\ell,L} - 1]$. We can assume $\tau_{c,k}^L < \tilde{\tau}_{k+1}^L = \tau_{k+1}^L$ and $\Pi_{\tau_{c,k}^L}^{\ell,L} = \tilde{\Pi}_{\tau_{c,k}^L}^{\ell,L}$ since these events have by the induction assumption asymptotically probability 1. Using the strong Markov property of $\Pi^{\ell,L}$ and $\tilde{\Pi}^{\ell,L}$ we can restart both processes at time $\tau_{c,k}^L$ (note that for $k + 1$ all statements only concern the processes at times greater than $\tau_{c,k}^L$). Since on the above events both processes start in the same partition (for which different blocks have different labels) the proof for $k + 1$ works now analogous to the case $k = 1$. The randomness of the new starting point poses no problem since we have uniform results (or alternatively because the space $\mathcal{P}_n^{\ell,L}$ is finite). \square

In order to study the asymptotic behavior of $\tilde{\Pi}^{\ell,L}$ and also of $\Pi^{\ell,L}$ we need some results on coalescing random walks that were already shown in Cox (1989)

Proposition 6.4. *Let Z^L be a simple symmetric random walk on T^L . Let $(a_L)_{L \in \mathbb{N}}$ be a sequence of nonnegative numbers with*

$$\lim_{L \rightarrow \infty} L^{-1} \sqrt{\log L} a_L = \infty, \quad (a_L)_{L \in \mathbb{N}} = o(L).$$

Let H^L be the first hitting time of the origin as in (41). Then

$$\lim_{L \rightarrow \infty} \sup_{\|z\| \geq a_L, t \geq 0} \left| \mathbb{P}^z \left(\frac{H^L}{s_L} > t \right) - \exp \left(-\frac{1}{2} \pi t \right) \right| = 0.$$

This is the special case $d = 2$ of Theorem 4 in Cox (1989).

Corollary 6.5. *Let $i \neq j \in [n]$. Then we have*

$$\lim_{L \rightarrow \infty} \sup_{\pi^{\ell,L} \in [[a_L, \sqrt{2}L]], i \approx_{\pi^{\ell,L}} j, t \geq 0} \left| \mathbb{P}^{\pi^{\ell,L}} \left(\frac{\tau^L(i, j)}{s_L} > t \right) - \exp(-\pi t) \right| = 0.$$

Proof. We know that $M_t^L(i) - M_t^L(j)$ (here the subtraction is done with respect to the cyclic structure of the torus) is up to time $\tau^L(i, j)$ a simple symmetric random walk on T^L with jump rate 2 and transition probabilities given by p^L that is started in $z := M_0^L(i) - M_0^L(j) \in T^L$ with $\|z\| \geq a_L$ due to $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$. The time $\tau^L(i, j)$ is just the hitting time of the origin of this random walk and so the claim follows from Proposition 6.4. \square

A more general statement holds for an arbitrary number of blocks that perform a coalescing random walk.

Proposition 6.6. *Let $\tilde{\tau}_1^L = \tau_1^L$ be the first meeting time of any blocks. Then for any $n \in \mathbb{N}$,*

$$\lim_{L \rightarrow \infty} \sup_{\pi^{\ell,L} \in [[a_L, \sqrt{2}L]] \cap \mathcal{P}_n^{\ell,L}, t \geq 0} \left| \mathbb{P}^{\pi^{\ell,L}} \left(\tau_1^L > s_L t \right) - \exp \left(-\pi \binom{\#\pi^{\ell,L}}{2} t \right) \right| = 0.$$

For $\#\pi^{\ell,L} = 2$ this is the statement of Corollary 6.5, for $\#\pi^{\ell,L} > 2$ see (3.2) in Cox (1989).

Lemma 6.7. *Let $n \in \mathbb{N}$ and let $K_C^L \subseteq \mathcal{P}_n^{\ell,L}$ for $C \subseteq [n]$ be those labeled partitions $\pi^{\ell,L}$, for which the blocks $\{A_0^L(i)\}_{i \in C}$ are pairwise different. Let $T > 0$ and $i, j, k, l \in [n]$ be pairwise disjoint. Then we have the following statements:*

(i) For $I_{\pi^{\ell,L}}^L := \int_0^{Ts_L} \mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) \in du, r_L(\tilde{M}_u^L(i), \tilde{M}_u^L(k)) \leq a_L)$ we have

$$\lim_{L \rightarrow \infty} \sup_{\pi^{\ell,L} \in [[a_L, \sqrt{2}L]] \cap K_{\{i,j,k\}}^L} I_{\pi^{\ell,L}}^L = 0.$$

(ii) For $J_{\pi^{\ell,L}}^L := \int_0^{Ts_L} \mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) \in du, r_L(\tilde{M}_u^L(k), \tilde{M}_u^L(l)) \leq a_L)$ we have

$$\lim_{L \rightarrow \infty} \sup_{\pi^{\ell,L} \in [[a_L, \sqrt{2}L]] \cap K_{\{i,j,k,l\}}^L} J_{\pi^{\ell,L}}^L = 0,$$

Proof. Since the structure of the blocks is irrelevant for the statement it suffices to consider $\tilde{\Pi}_0^{\ell,L} = \pi^{\ell,L} = \{\{i\} \mid i \in [n]\}$. The two statements for the coalescing random walk $\tilde{\Pi}^{\ell,L}$ are now (3.7) und (3.8) in Cox (1989). \square

The next result states that the probability for the event $B^L(q, \pi^{\ell,L})$ of (44) tends to 1 as L is large and q is small, provided the blocks are sampled far enough apart.

Proposition 6.8. *For any $\varepsilon > 0$ there is a $q(\varepsilon) > 0$ and $L(\varepsilon) \in \mathbb{N}$ such that for all $L \geq L(\varepsilon)$ and all $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ we have*

$$\mathbb{P}^{\pi^{\ell,L}}(B^L(q(\varepsilon), \pi^{\ell,L})) \geq 1 - \varepsilon.$$

Proof. We will show the following statements by induction over $k \in [\#\pi^{\ell,L} - 1]$ for any $\varepsilon > 0$. There is a $q(\varepsilon) > 0$ and $L(\varepsilon) \in \mathbb{N}$ such that for all $L \geq L(\varepsilon)$ and all $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with $\#\pi^{\ell,L} \geq k$ we have

$$\mathbb{P}^{\pi^{\ell,L}}(\tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L} \in [[a_L, \sqrt{2}L]]) \geq 1 - \varepsilon, \quad (53)$$

$$\mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_k^L - \tilde{\tau}_{k-1}^L > s_L q(\varepsilon)) \geq 1 - \varepsilon. \quad (54)$$

In words, these events say that at the time $\tilde{\tau}_k^L$, the blocks are again far apart with high probability and the probability of meeting times being very close together is small. The proposition follows then from (54) for all $k \in [\#\pi^{\ell,L} - 1]$ since $\tilde{\tau}_k^L - \tilde{\tau}_{k-1}^L$ and $\tilde{\tau}_k^{L,(2)} - \tilde{\tau}_{k-1}^L$ have the same distribution.

Note that since in $\tilde{\Pi}^{\ell,L}$ almost surely only pairwise mergers occur, all of the $\tilde{\tau}_k^L$ above are actual jump times of the process.

We start with $k = 1$ and first show (53). From Proposition 6.6 it follows that there is an $L(\varepsilon) \in \mathbb{N}$ and a $T(\varepsilon) > 0$ such that

$$\mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L > s_L T(\varepsilon)) < \frac{\varepsilon}{2}$$

for all $L \geq L(\varepsilon)$ and all $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with $\#\pi^{\ell,L} \geq 2$. According to Lemma 6.7, by choosing $L(\varepsilon)$ larger if necessary,

$$\mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L \in [0, s_L T(\varepsilon)], \tilde{\Pi}_{\tilde{\tau}_1^L}^{\ell,L} \notin [[a_L, \sqrt{2}L]]) < \frac{\varepsilon}{2}$$

for all $L \geq L(\varepsilon)$ and all $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with $\#\pi^{\ell,L} \geq 2$. Thus, for those L and $\pi^{\ell,L}$

$$\begin{aligned} & \mathbb{P}^{\pi^{\ell,L}}(\tilde{\Pi}_{\tilde{\tau}_1^L}^{\ell,L} \notin [[a_L, \sqrt{2}L]]) \\ & \leq \mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L \in [0, s_L T(\varepsilon)], \tilde{\Pi}_{\tilde{\tau}_1^L}^{\ell,L} \notin [[a_L, \sqrt{2}L]]) + \mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L > s_L T(\varepsilon)) < \varepsilon. \end{aligned}$$

In order to show (54) we note that by Corollary 6.5 we can choose $q(\varepsilon) > 0$ small enough and make $L(\varepsilon) \in \mathbb{N}$ larger if necessary such that for all $L \geq L(\varepsilon)$ and all $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with $\#\pi^{\ell,L} \geq 2$

$$\left| \mathbb{P}^{\pi^{\ell,L}}(s_L^{-1} \tau^L(i, j) \leq q(\varepsilon)) \right| < \frac{\varepsilon}{n^2}. \quad (55)$$

With $C^L(0)$ as in (47) it follows that

$$\begin{aligned} \mathbb{P}^{\pi^{\ell,L}}(\tilde{\tau}_1^L - \tilde{\tau}_0^L \leq s_L q(\varepsilon)) & \leq \sum_{(i,j) \in C^L(0)} \mathbb{P}^{\pi^{\ell,L}}(s_L^{-1} \tilde{\tau}_1^L = s_L^{-1} \tau^L(i, j) \leq q(\varepsilon)) \\ & \leq \sum_{(i,j) \in C^L(0)} \mathbb{P}^{\pi^{\ell,L}}(s_L^{-1} \tau^L(i, j) \leq q(\varepsilon)) < \sum_{(i,j) \in C^L(0)} \frac{\varepsilon}{n^2} \leq \varepsilon. \end{aligned}$$

Let the claim now be true for $k \in [n-2]$. Since (53) holds for k we can assume

$$\tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L} \in [[a_L, \sqrt{2}L]].$$

Using the strong Markov property of $\tilde{\Pi}^{\ell,L}$ we can restart the process at time $\tilde{\tau}_k^L$. Since the blocks are again well separated the induction step now follows analogous to the $k=1$ proof. Note that the random starting point does not pose a problem since we have uniform results (or alternatively because \mathcal{P}_n^L is finite). \square

We will need the following properties of a random walk on the torus, which states that asymptotically the random walk will be uniformly distributed.

Proposition 6.9. *Let Z^L be a simple symmetric random walk on T^L with transition kernel p^L . Let $(t_L)_{L \in \mathbb{N}}$ be a sequence with $\lim_{L \rightarrow \infty} t_L = \infty$. Then*

$$\lim_{L \rightarrow \infty} \sup_{t \geq t_L} \sup_{(2L+1)^2} (2L+1)^2 |p_t^L(x, 0) - (2L+1)^{-2}| = 0$$

This is (2.8) in Cox (1989). From Proposition 6.9 we obtain the asymptotic exchangeability of the blocks of the coalescing random walk.

Lemma 6.10. *Let $\pi_0 \in \mathcal{P}_n$ be a partition with $\#\pi_0 \geq 2$ and for $L \in \mathbb{N}$ large enough let $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ be a labeled partition, whose partition structure equals π_0 , meaning that $\pi^L = \pi_0$. Furthermore, let σ be a permutation of $[\#\pi_0]$ and $\pi^{\sigma,\ell,L}$ the labeled partition that is obtained from $\pi^{\ell,L}$ by permuting the labels of the block with σ . We define for $L \in \mathbb{N}$*

$$q_L := (\log \log L)(2L+1)^2.$$

Let $k \in [\#\pi_0 - 1]$, then there exists a sequence $(\delta_L)_{L \in \mathbb{N}}$, independent of $\pi^{\ell,L}$ and σ , with $\delta_L \xrightarrow{L \rightarrow \infty} 0$ and

$$\left| \mathbb{E}^{\pi^{\ell,L}}(f(\tilde{\Pi}_{q_L}^{\ell,L})) - \mathbb{E}^{\pi^{\sigma,\ell,L}}(f(\tilde{\Pi}_{q_L}^{\ell,L})) \right| < \delta_L,$$

for any measurable and bounded $f: \mathcal{P}_n^{\ell,L} \rightarrow \mathbb{R}$.

Proof. Let $m := \#\pi_0$ and $\pi^{\ell,L} = \{(B_1, \zeta_1), \dots, (B_m, \zeta_m), \dots\}$. We can couple the coalescing random walks started in $\pi^{\ell,L}$ and $\pi^{\sigma,\ell,L}$ in a natural way by using the same motions for the corresponding blocks that start with the same labels. Thus, all times $\tilde{\tau}_k^L = \tilde{\tau}_{c,k}^L$ are identical for $k = 1, \dots, m-1$. Since $s_L^{-1} q_L \rightarrow 0$ as $L \rightarrow \infty$ we have by Proposition 6.6 that

$$\mathbb{P}^{\pi^{\ell,L}}(q_L < \tilde{\tau}_1^L) = \mathbb{P}^{\pi^{\sigma,\ell,L}}(q_L < \tilde{\tau}_1^L) \xrightarrow{L \rightarrow \infty} 1 \quad (56)$$

uniformly over all $\pi^{\ell,L}, \pi^{\sigma,\ell,L} \in [[a_L, \sqrt{2}L]]$. Thus, due to the boundedness of f it suffices to show that claim on the event $\{q_L < \tilde{\tau}_1^L < \infty\}$. For any $z = (z_1, \dots, z_m) \in (T^L)^m$ we set $\pi_z = ((B_1, z_1), \dots, (B_m, z_m), (\emptyset, \partial), \dots)$. While $\{t \leq q_L < \tilde{\tau}_1^L\}$ the blocks perform independent random walks with transition kernel p^L . Thus, we have due to (56) that

$$\mathbb{E}^{\pi^{\ell,L}}(f(\tilde{\Pi}_{q_L}^{\ell,L})) = \sum_{z \in (T^L)^m} f(\pi_z) \mathbb{P}^{\pi^{\ell,L}}(\tilde{\Pi}_{q_L}^{\ell,L} = \pi_z) + o(1) = \sum_{z \in (T^L)^m} f(\pi_z) \prod_{k=1}^m p_{q_L}^L(\zeta_k - z_k, 0) + o(1),$$

where the $o(1)$ term converges to 0 uniformly for $L \rightarrow \infty$. We also have the analogous statement for $\pi^{\sigma,\ell,L}$ with $\sigma(\zeta_k)$ instead of ζ_k . Hence,

$$\begin{aligned} & \left| \mathbb{E}^{\pi^{\ell,L}}(f(\tilde{\Pi}_{q_L}^{\ell,L})) - \mathbb{E}^{\pi^{\sigma,\ell,L}}(f(\tilde{\Pi}_{q_L}^{\ell,L})) \right| \\ & \leq \sum_{z \in (T^L)^m} |f(\pi_z)| \left| \left(\prod_{k=1}^m p_{q_L}^L(\zeta_k - z_k, 0) - \prod_{k=1}^m p_{q_L}^L(\sigma(\zeta_k) - z_k, 0) \right) \right| + o(1) \\ & \leq \sum_{z \in (T^L)^m} \left| f(\pi_z) \left(\prod_{k=1}^m p_{q_L}^L(\zeta_k - z_k, 0) \right) - (2L+1)^{-2m} \right| \\ & \quad + \sum_{z \in (T^L)^m} \left| f(\pi_z) \left(\prod_{k=1}^m p_{q_L}^L(\sigma(\zeta_k) - z_k, 0) \right) - (2L+1)^{-2m} \right| + o(1) \\ & \leq 2 \|f\|_\infty \sup_{z \in (T^L)^m} \left| \left(\prod_{k=1}^m (2L+1)^2 p_{q_L}^L(z_k, 0) \right) - 1 \right| + o(1), \end{aligned}$$

where we have used in the last inequality that there are $(2L+1)^{2m}$ terms in the sums of the previous line. With Proposition 6.9 (set $t_L := \log \log L$) we have $(2L+1)^2 p_{q_L}^L(z, 0) \rightarrow 1$ uniformly in z and so the claim now follows since the right hand side converges uniformly over all $\pi^{\ell,L}$ and $\pi^{\sigma,\ell,L}$. \square

We now introduce some notation for the waiting time between meeting and coalescence times of blocks. Namely, for $\pi^{\ell,L} \in \mathcal{P}_n^{\ell,L}$ we define for $k \in [\#\pi^L - 1]$ the *waiting times*

$$\sigma_k^L := \tau_k^L - \tau_{k-1}^L, \quad \sigma_{c,k}^L := \tau_{c,k}^L - \tau_{c,k-1}^L. \quad (57)$$

(Formally, set $\infty - \infty = 0$.) Analogously we define the waiting times $\tilde{\sigma}_k^L$ of $\tilde{\Pi}_k^{\ell,L}$. For $\pi_0 \in \mathcal{P}_n$ let $(K_t^{\pi_0})_{t \in \mathbb{R}_+}$ be the non-spatial Kingman coalescent started in π_0 and let $\tau_{K,k}$ be its k -th coalescence time for $k \in [\#\pi_0 - 1]$. We set $\tau_{K,0} := 0$ and define for $k \in [\#\pi - 1]$ the waiting times

$$U_k := \tau_{K,k} - \tau_{K,k-1}. \quad (58)$$

Note that the family of random variables $\{U_k\}_{k \in [\#\pi_0 - 1]}$ is independent, and that

$$U_k \sim \text{Exp}\left(\binom{\#\pi_0 - (k - 1)}{2}\right).$$

Note also that they are independent of $\{K_{\tau_{K,k}}^{\pi_0}\}_{k \in [\#\pi_0 - 1]}$ as well as that $K_{\tau_{K,k}}^{\pi_0}$ results from $K_{\tau_{K,k-1}}^{\pi_0}$ by coalescence of two blocks chosen at random.

We will now be able to prove a result about the asymptotic behavior of the coalescence times and the types of transitions for coalescing random walks. Note that the result refers only to the partition structure $\tilde{\Pi}^L$ of the coalescing random walk and not to the labeled partitions $\tilde{\Pi}^{\ell,L}$.

Theorem 6.11. *Let $\pi_0 \in \mathcal{P}_n$ with $m := \#\pi_0 \geq 2$. We consider for every $L \in \mathbb{N}$ large enough a labeled partition $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with partition structure $\pi^L = \pi_0$. We then obtain convergence in distribution for $L \rightarrow \infty$,*

$$\left(\frac{\tilde{\sigma}_1^L}{s_L}, \tilde{\Pi}_{\tilde{\tau}_1^L}^L, \dots, \frac{\tilde{\sigma}_{m-1}^L}{s_L}, \tilde{\Pi}_{\tilde{\tau}_{m-1}^L}^L\right) \Rightarrow (\pi U_1, K_{\tau_{K,1}}^{\pi_0}, \dots, \pi U_{m-1}, K_{\tau_{K,m-1}}^{\pi_0}).$$

Proof. The proof will follow along the lines of the analogous result in Limic and Sturm (2006) for (U_1, \dots, U_{m-1}) in dimension $d \geq 3$ and also use the results of Cox (1989) for coalescing random walks. We first show that there is a sequence $(\varepsilon_L)_{L \in \mathbb{N}}$ with $\varepsilon_L \xrightarrow{L \rightarrow \infty} 0$ such that for all $k \in [m - 1]$ and $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ as well as for all $u \geq 0$,

$$\mathbb{E}^{\pi^{\ell,L}} \left(\left| \mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L}}(\tilde{\tau}_1^L > s_L u) - \exp\left(-\pi \binom{\#\pi - k}{2} u\right) \right| \right) < \varepsilon_L. \quad (59)$$

Note that since no random walk jumps happen at the same time and since coalescence is instantaneous, no more than two blocks may meet and coalesce, so that $\#\tilde{\Pi}_{\tilde{\tau}_k^L}^L = \#\pi - k$ for all $k \in [m - 1]$. For those k we also set

$$D_k^L := \left\{ \tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L} \in [[a_L, \sqrt{2}L]] \right\}.$$

Then, for all $u \geq 0$ we obtain

$$\begin{aligned} & \mathbb{E}^{\pi^{\ell,L}} \left(\left| \mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L}}(\tilde{\tau}_1^L > s_L u) - \exp\left(-\pi \binom{\#\pi - k}{2} u\right) \right| \right) \\ & \leq \mathbb{E}^{\pi^{\ell,L}} \left(1_{D_k^L} \left| \mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_k^L}^{\ell,L}}(\tilde{\tau}_1^L > s_L u) - \exp\left(-\pi \binom{\#\pi - k}{2} u\right) \right| \right) + 2\mathbb{P}^{\pi^{\ell,L}}((D_k^L)^C). \end{aligned} \quad (60)$$

Proposition 6.6 and Lebesgue's dominated convergence theorem imply that the first term on the right hand side converges to zero uniformly in $\pi^{\ell,L}$ and u . The second term on the right hand side of (60) converges to 0 uniformly in $\pi^{\ell,L}$ and u due Proposition 6.8 (see (53) in the proof). This shows (59).

We will now show convergence in distribution. For $t_1, \dots, t_{m-1} \geq 0, \pi_1, \dots, \pi_{m-1} \in \mathcal{P}_n$ and $k \in [m - 1]$ we define the event

$$A_k^L := \left\{ \frac{\tilde{\sigma}_k^L}{s_L} > t_k, \tilde{\Pi}_{\tilde{\tau}_k^L}^L = \pi_k, \dots, \frac{\tilde{\sigma}_1^L}{s_L} > t_1, \tilde{\Pi}_{\tilde{\tau}_1^L}^L = \pi_1 \right\}.$$

Also set $A_0^L := \Omega$. Since only binary mergers are possible for coalescing random walks as well as the Kingman coalescent, it suffices to consider π_k chosen such that π_k results from π_{k-1} through a coalescence of exactly two blocks for all $k \in [m-1]$. We define

$$C_k := \{(\min B_1, \min B_2) \mid B_1, B_2 \text{ are blocks of } \pi_k, \min B_1 < \min B_2\}.$$

Thus, we have $|C_k| = \binom{\#\pi_k}{2} = \binom{m-k}{2}$. Let $(i, j) \in C_{k-1} \setminus C_k$ meaning that on the event A_k^L at time $\tilde{\tau}_k^L$ the blocks containing i and j in the partition π_{k-1} merge to form the partition π_k . Thus, on the event A_{k-1}^L we have $\{\tilde{\Pi}_{\tilde{\tau}_k^L}^L = \pi_k\} = \{\tilde{\tau}_k^L = \tilde{\tau}^L(i, j)\}$. We set $q_L := (\log \log L)(2L+1)^2$ for all $L \in \mathbb{N}$ and assume that we consider an L large enough so that $q_L < s_L t_k$ for $k \in [m-1]$. Let $\tilde{\mathcal{F}}_t^L$ be the σ -algebra generated by $\tilde{\Pi}^{\ell, L}$ up to time t . Then A_{k-1}^L is $\tilde{\mathcal{F}}_{\tilde{\tau}_{k-1}^L}^L$ measurable and we have that

$$\begin{aligned} \mathbb{P}^{\pi^{\ell, L}}(A_k^L) &= \mathbb{E}^{\pi^{\ell, L}}(\mathbb{E}^{\pi^{\ell, L}}(1_{\{\tilde{\sigma}_k^L > s_L t_k, \tilde{\Pi}_{\tilde{\tau}_k^L}^L = \pi_k\}} \cdot 1_{A_{k-1}^L} | \tilde{\mathcal{F}}_{\tilde{\tau}_{k-1}^L}^L)) \\ &= \mathbb{E}^{\pi^{\ell, L}}\left(\mathbb{P}^{\pi^{\ell, L}}\left(\tilde{\sigma}_k^L > s_L t_k, \tilde{\tau}_k^L = \tilde{\tau}^L(i, j) \mid \tilde{\mathcal{F}}_{\tilde{\tau}_{k-1}^L}^L\right) 1_{A_{k-1}^L}\right) \\ &= \mathbb{E}^{\pi^{\ell, L}}\left(\mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^L}\left(\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) > s_L t_k\right) 1_{A_{k-1}^L}\right). \end{aligned} \quad (61)$$

By using that $\{\tilde{\tau}_1^L > s_L t_k\} = \{\tilde{\tau}_1^L > q_L, \tilde{\tau}_1^L - q_L > s_L t_k - q_L\}$ we obtain from conditioning on the information up to time q_L that

$$\mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^L}\left(\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) > s_L t_k\right) = \mathbb{E}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^L}\left(f\left(\tilde{\Pi}_{q_L}^{\ell, L}\right) 1_{\{\tilde{\tau}_1^L > q_L\}}\right), \quad (62)$$

where the function f is defined by

$$f: \mathcal{P}_n^{\ell, L} \rightarrow \mathbb{R}, \quad \pi \mapsto \mathbb{P}^\pi\left(\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) > s_L t_k - q_L\right).$$

Note that f is measurable and bounded. For $(l, m) \in C_{k-1}$ let $\sigma_{l, m}$ be the permutation of $[\#\pi_{k-1}]$ that only swaps i with l and also j with m . Let $\pi^{\sigma_{l, m}, \ell, L}$ be the partition obtained from $\pi^{\ell, L}$ by permuting the labels with $\sigma^{l, m}$ (as in Lemma 6.10). We also set $\hat{A}_k^L := A_k^L \cap D_k^L$. Then, due to (61) and (62),

$$\begin{aligned} & \left| \mathbb{P}^{\pi^{\ell, L}}(A_k^L) - \mathbb{E}^{\pi^{\ell, L}}\left(\sum_{(l, m) \in C_{k-1}} \frac{1}{|C_{k-1}|} \mathbb{E}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\sigma_{l, m}, \ell, L}}\left(f\left(\tilde{\Pi}_{q_L}^{\ell, L}\right) 1_{\{\tilde{\tau}_1^L > q_L\}}\right) 1_{A_{k-1}^L}\right) \right| \\ & \leq \mathbb{E}^{\pi^{\ell, L}}\left(\sum_{(l, m) \in C_{k-1}} \frac{1}{|C_{k-1}|} \left| \mathbb{E}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\ell, L}}\left(f\left(\tilde{\Pi}_{q_L}^{\ell, L}\right) 1_{\{\tilde{\tau}_1^L > q_L\}}\right) - \mathbb{E}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\sigma_{l, m}, \ell, L}}\left(f\left(\tilde{\Pi}_{q_L}^{\ell, L}\right) 1_{\{\tilde{\tau}_1^L > q_L\}}\right) \right| 1_{\hat{A}_{k-1}^L}\right) + 2\mathbb{P}^{\pi^{\ell, L}}((D_{k-1}^L)^C). \end{aligned}$$

Since due to (59) the probability for the event $\{\tilde{\tau}_1^L > q_L\}$ is arbitrarily close to 1 for large L , we have that the first term on the right hand side converges to 0 as $L \rightarrow \infty$ because of Lemma 6.10 and Lebesgue's dominated convergence theorem. The second term converges to 0 due to Proposition 6.8. Thus, we obtain

$$\mathbb{P}^{\pi^{\ell, L}}(A_k^L) = |C_{k-1}|^{-1} \mathbb{E}^{\pi^{\ell, L}}\left(\sum_{(l, m) \in C_{k-1}} \mathbb{E}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\sigma_{l, m}, \ell, L}}\left(1_{\{\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) > s_L t_k\}}\right) 1_{A_{k-1}^L}\right) + o(1).$$

Note that for the process started in $\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\sigma_{l,m,\ell,L}}$ the event $\{\tilde{\tau}_1^L = \tilde{\tau}^L(i, j) > s_L t_k\}$ is the same as the event $\{\tilde{\tau}_1^L = \tilde{\tau}^L(l, m) > s_L t_k\}$ for the process started in $\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\ell,L}$. Since the events $\{\tilde{\tau}^L(l, m) = \tilde{\tau}_1^L, (l, m) \in C_{k-1}\}$ are a partition of the probability space it then follows that

$$\begin{aligned}\mathbb{P}^{\pi^{\ell,L}}(A_k^L) &= |C_{k-1}|^{-1} \mathbb{E}^{\pi^{\ell,L}} \left(\mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\ell,L}} \left(\tilde{\tau}_1^L > s_L t_k \right) 1_{A_{k-1}^L} \right) + o(1) \\ &= |C_{k-1}|^{-1} \mathbb{E}^{\pi^{\ell,L}} \left(\left(\mathbb{P}^{\tilde{\Pi}_{\tilde{\tau}_{k-1}^L}^{\ell,L}} \left(\tilde{\tau}_1^L > s_L t_k \right) - \exp \left(-\pi \binom{m-(k-1)}{2} t_k \right) \right) 1_{A_{k-1}^L} \right) \\ &\quad + |C_{k-1}|^{-1} \exp \left(-\pi \binom{m-(k-1)}{2} t_k \right) \mathbb{P}^{\pi^{\ell,L}}(A_{k-1}^L) + o(1).\end{aligned}$$

It now follows from (59) that

$$\lim_{L \rightarrow \infty} \mathbb{P}^{\pi^{\ell,L}}(A_k^L) = \binom{m-(k-1)}{2}^{-1} \exp \left(-\pi \binom{m-(k-1)}{2} t_k \right) \lim_{L \rightarrow \infty} \mathbb{P}^{\pi^{\ell,L}}(A_{k-1}^L).$$

By induction we then have that

$$\begin{aligned}\lim_{L \rightarrow \infty} \mathbb{P}^{\pi^{\ell,L}}(A_k^L) &= \lim_{L \rightarrow \infty} \prod_{i=1}^k \binom{m-(i-1)}{2}^{-1} \exp \left(-\pi \binom{m-(i-1)}{2} t_i \right) \\ &= \prod_{i=1}^k \mathbb{P}(K_{\tau_{K,i}}^{\pi_0} = \pi_i | K_{\tau_{K,i-1}}^{\pi_0} = \pi_{i-1}) \cdot \mathbb{P}(\pi U_i > t_i).\end{aligned}$$

Since this is the desired quantity this finishes the proof of convergence in distribution. \square

We now formulate the corresponding result for Λ - n -coalescents.

Corollary 6.12. *Let $\pi_0 \in \mathcal{P}_n$ with $m = \#\pi_0$. We consider for $L \in \mathbb{N}$ large enough a labeled partition $\pi^{\ell,L} \in [[a_L, \sqrt{2}L]]$ with corresponding partition structure $\pi^L = \pi_0$. Then there is convergence in distribution for $L \rightarrow \infty$,*

$$\left(\frac{\sigma_{c,1}^L}{s_L}, \Pi_{\tau_{c,1}^L}^L, \dots, \frac{\sigma_{c,m-1}^L}{s_L}, \Pi_{\tau_{c,m-1}^L}^L \right) \Rightarrow (\pi U_1, K_{\tau_{K,1}}^{\pi_0}, \dots, \pi U_{m-1}, K_{\tau_{K,m-1}}^{\pi_0}).$$

Proof. For $L \in \mathbb{N}$ we set

$$\begin{aligned}V^L &:= \left(\frac{\tilde{\sigma}_1^L}{s_L}, \tilde{\Pi}_{\tilde{\tau}_1^L}^L, \dots, \frac{\tilde{\sigma}_{m-1}^L}{s_L}, \tilde{\Pi}_{\tilde{\tau}_{m-1}^L}^L \right), \\ V_c^L &:= \left(\frac{\sigma_{c,1}^L}{s_L}, \Pi_{\tau_{c,1}^L}^L, \dots, \frac{\sigma_{c,m-1}^L}{s_L}, \Pi_{\tau_{c,m-1}^L}^L \right), \\ V_K &:= (\pi U_1, K_{\tau_{K,1}}^{\pi_0}, \dots, \pi U_{m-1}, K_{\tau_{K,m-1}}^{\pi_0}).\end{aligned}$$

Due to Propositions 6.3 and 6.8 we have that $V_c^L - V^L$ converges to 0 in probability. From Theorem 6.11 we obtain that $V^L \Rightarrow V_K$. Taken together this implies that $V_c^L \Rightarrow V_K$ as required. \square

We are finally ready to prove our main result, Theorem 4.1.

Proof. For any fixed n we prove convergence in the Skorohod space $D(\mathbb{R}_+, \mathcal{P}_n)$,

$$(\Pi_{s_L t}^L)_{t \in \mathbb{R}_+} \Rightarrow (K_{\pi t}^{\pi_0})_{t \in \mathbb{R}_+},$$

by first showing relative compactness and then weak convergence of the finite dimensional distributions, see Theorem 3.7.8 of Ethier and Kurtz (1986). We start with the relative compactness of $((\Pi_{s_L t}^L)_{t \in \mathbb{R}_+})_{L \in \mathbb{N}}$. Since \mathcal{P}_n is compact it suffices to show by Theorem 3.6.3 of Ethier and Kurtz (1986) that for all $T > 0$ and $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\limsup_{L \rightarrow \infty} \mathbb{P}^{\pi^{f,L}}(w(\Pi_{s_L \cdot}^L, \delta, T) \geq \varepsilon) \leq \varepsilon. \quad (63)$$

where $w(\Pi_{s_L \cdot}^L, \delta, T)$ is the δ -modulus of continuity of $\Pi_{s_L \cdot}^L$ on $[0, T]$ (see also (37)). Namely, it is the infimum over all partitions of the form $0 = t_0 < t_1 < \dots < t_{k-1} < T \leq t_k$ such that $t_i - t_{i-1} > \delta$ for all $1 \leq i \leq k$ of the quantity

$$\max_i \sup_{r, t \in [t_{i-1}, t_i)} 1 \vee d_n(\Pi_{s_L r}^L, \Pi_{s_L t}^L).$$

We set $T > 0$ and $\delta > 0$ as well as

$$c_1 := \max \left\{ \pi \binom{\#\pi_0 - (k-1)}{2} \middle| k \in [\#\pi_0 - 1] \right\}.$$

Due to Corollary 6.12 there is an $L(\delta)$ such that for all $L \geq L(\delta)$ and all $k \in [\#\pi_0 - 1]$ we have

$$\begin{aligned} \mathbb{P}^{\pi^{f,L}}(s_L^{-1} \sigma_{c,k}^L < 2\delta \text{ for one } k \in [\#\pi_0 - 1]) &< \delta + \mathbb{P}^{\pi_0}(\pi U_k \leq 2\delta \text{ for one } k \in [\#\pi_0 - 1]) \\ &\leq \delta + (\#\pi_0 - 1)(1 - \exp(-2c_1\delta)) \leq (1 + 2c_1(\#\pi_0 - 1))\delta. \end{aligned}$$

Since $(\Pi_{s_L t}^L)_{t \in \mathbb{R}_+}$ is constant on the intervals $[s_L^{-1} \tau_{c,k-1}^L, s_L^{-1} \tau_{c,k}^L)$ we have on the event that $s_L^{-1} \sigma_{c,k}^L > 2\delta$ for all $k \in [\#\pi_0 - 1]$ that $w(\Pi_{s_L \cdot}^L, \delta, T) = 0$. Hence, (63) and so the relative compactness follows by choosing $\delta = \varepsilon(1 + 2c_1(\#\pi_0 - 1))^{-1}$.

Lastly, we show the weak convergence of the finite dimensional distributions. Let $l \in \mathbb{N}$ and $t_1, \dots, t_l \in \mathbb{R}_+$. By definition

$$\Pi_{s_L t}^L = \sum_{k=1}^{\#\pi_0 - 1} 1_{\{s_L t \in [\tau_{c,k-1}^L, \tau_{c,k}^L)\}} \Pi_{\tau_{c,k-1}^L}^L. \quad (64)$$

We define for $k \in [l]$ the function $f_k: (\mathbb{R}_+ \times \mathcal{P}_n)^{\#\pi_0 - 1} \rightarrow \mathcal{P}_n$ by

$$(r_1, \pi_1, \dots, r_{\#\pi_0 - 1}, \pi_{\#\pi_0 - 1}) \mapsto \begin{cases} \pi_0 & , t_k \in [0, r_1) \\ \pi_i & , t_k \in \left[\sum_{j=1}^i r_j, \sum_{j=1}^{i+1} r_j \right), i < \#\pi_0 - 1 \\ \pi_{\#\pi_0 - 1} & , t_k \in \left[\sum_{j=1}^{\#\pi_0 - 1} r_j, \infty \right). \end{cases}$$

We define the function

$$f := (f_1, \dots, f_l): (\mathbb{R}_+ \times \mathcal{P}_n)^{\#\pi_0 - 1} \rightarrow (\mathcal{P}_n)^l.$$

Let V_c^L and V_K be defined as in the proof of Corollary 6.12. Due to (64) we have

$$f(V_c^L) = (\Pi_{s_L t_1}^L, \dots, \Pi_{s_L t_l}^L),$$

and likewise

$$f(V_K) = (K_{\pi_{t_1}}^{\pi_0}, \dots, K_{\pi_{t_l}}^{\pi_0}).$$

Since U_k are continuous random variables we note that the event that V_K takes values in the discontinuity set of f has probability 0. Thus, due to $V_c^L \Rightarrow V_K$ from Corollary 6.12 we obtain

$$(\Pi_{s_{L t_1}}^L, \dots, \Pi_{s_{L t_l}}^L) = f(V_c^L) \Rightarrow f(V_K) = (K_{\pi_{t_1}}^{\pi_0}, \dots, K_{\pi_{t_l}}^{\pi_0}).$$

This finishes the proof. \square

References

- [1] Barton, N., Etheridge, A., Véber, A., 2010. A new model for evolution in a spatial continuum. *Electron. J. Probab.* 15, 162–216.
- [2] Berestycki, J., Berestycki, N., Limic, V., unpublished. Asymptotic sampling formulae for Λ -coalescents. Preprint at arxiv:1201.6512, 1–20.
- [3] Berestycki, J., Berestycki, N., Schweinsberg, J., 2007. Beta-coalescents and continuous stable random trees. *Ann. Probab.* 35, 1835–1887.
- [4] Birkner, M., Blath, J., Steinrück, M., in press 2011. Importance sampling for Λ -coalescents in the infinitely many sites model. *Theor. Pop. Biol.*, 1–32.
- [5] Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: A new approach 1. haploid models. *Adv. Appl. Probab.* 6, 260–290.
- [6] Cox, J., 1989. Coalescing random walks and voter model consensus times on the torus in \mathbf{z}^d . *Ann. Probab.* 17, 1333–1366.
- [7] Cox, J., Durrett, R., 2002. The stepping stone model: New formulas expose old myths. *Ann. Appl. Probab.* 12, 1348–1377.
- [8] Donnelly, P., Kurtz, T., 1999. Particle representations for measure-valued population models. *Ann. Probab.* 27, 166–205.
- [9] Durrett, R., Schweinsberg, J., 2004. Approximating selective sweeps. *Theor. Pop. Biol.* 66, 129–138.
- [10] Durrett, R., Schweinsberg, J., 2005. Random partitions approximating the coalescence of lineages during a selective sweeps. *Ann. Appl. Probab.* 15, 1591–1651.
- [11] Eldon, B., Wakeley, J., 2006. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* 172, 2621–2633.
- [12] Etheridge, A., Véber, A., in press 2012. The spatial Fleming-Viot process on a large torus: genealogies in the presence of recombination. To appear in *Ann. Appl. Probab.*, Preprint at arXiv:1106.4050, 1–35.
- [13] Ethier, S.N., Kurtz, T.G., 1986. *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics, Wiley.
- [14] Greven, A., Limic, V., Winter, A., 2005. Representation theorems for interacting Moran models, interacting Fisher-Wright diffusions and applications. *Electron. J. Probab.* 10, 1286–1358.
- [15] Greven, A., Limic, V., Winter, A., unpublished. Coalescent processes arising in a study of diffusive clustering. Preprint at arXiv:math.PR/0703875.
- [16] Herbots, H.M., 1997. The structured coalescent, in: Donnelly, P., Tavaré, S. (Eds.), *Progress of Population Genetics and Human Evolution*. Springer, pp. 231–255.
- [17] Kingman, J.F., 1982a. The coalescent. *Stoch. Proc. Appl.* 13, 249–261.
- [18] Kingman, J.F., 1982b. On the genealogy of large populations, in: Special volume of *Journal of Applied Probability*. *Journal of Applied Probability*. volume 19A, pp. 27–43.
- [19] Limic, V., Sturm, A., 2006. The spatial Λ -coalescent. *Electron. J. Probab.* 11, 363–393.
- [20] Möhle, M., 2006a. On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* 12, 35–53.
- [21] Möhle, M., 2006b. On the number of segregating sites for populations with large family sizes. *Adv. Appl. Probab.* 38, 750–767.
- [22] Möhle, M., Sagitov, S., 2001. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* 29, 1547–1562.
- [23] Notohara, M., 1990. The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.* 31, 841–852.
- [24] Pitman, J., 1999. Coalescents with multiple collisions. *Ann. Probab.* 27, 1870–1902.

- [25] Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* 36, 1116–1125.
- [26] Wakeley, J., 2001. The coalescent in an island model of population subdivision with variation among demes. *Theor. Pop. Biol.* 59, 133–144.
- [27] Wakeley, J., 2009. *Coalescent Theory: An Introduction*. Roberts & Company.
- [28] Zähle, I., Cox, J., Durrett, R., 2005. The stepping stone model II: Genealogies and the infinite sites model. *Ann. Appl. Probab.* 15(1B), 671–699.